2019

Essays On The Economics Of Networks

Yiran Chen
University of Pennsylvania, yiranchen.econ@gmail.com

Essays On The Economics Of Networks

Abstract

Empirical studies of social and economic networks are facilitated by the growing availability of network data. This area of research focuses on understanding two major questions: how networks affect economic outcomes and how networks are formed. This dissertation studies these questions respectively. The first chapter examines the impact of social networks on agents' economic outcomes in the context of job referrals in the labor market. The second chapter relates to the formation of financial networks with latent traits in the context of U.S. campaign contributions.

In the first chapter, "A Structural Analysis of Job Referrals and Social Networks: The Case of the Corporate Executives Market", I develop and structurally implement a labor market search model in which workers, in addition to directly receiving job offers, also receive referrals from their social contacts. In the model, referrals are generated endogenously: an external referral occurs when a friend rejects an offer he/she receives, and an internal referral occurs when a friend leaves his/her current job. I estimate the model by Generalized Method of Moments using data on the labor market history and the social connections of executives in S&P 500 firms. Using the estimated model, I find that referrals play a substantial role in the executive labor market. More than one quarter of the job transitions and raises are driven byvreferrals. Shutting down referrals reduces executives' welfare by an equivalence of a two to seven percentage points reduction in income. I also evaluate the impacts of the social networks' structure by comparing the outcomes under the observed networks and alternative randomly formed networks. I find that the welfare distribution is more unequal under the random networks. I further investigate the mechanisms for these effects through the lens of two network statistics: friends' popularity and local community clustering.

In the second chapter, "Inferring the Ideological Affiliations of Political Committees via Financial Contributions Networks" (co-authored with Hanming Fang), we address the missing data problem for about two thirds of the political committees that do not self-identify their party affiliations in their registration with the Federal Election Commission. In this chapter, we propose and implement a novel Bayesian approach to infer the ideological affiliations of political committees based on the network of financial contributions among them. In Monte Carlo simulations, we demonstrate that our estimation algorithm achieves very high accuracy in recovering these committees' latent ideological affiliations when the pairwise difference in ideology groups' connection patterns satisfy a condition known as the Chernoff-Hellinger divergence criterion. We illustrate our approach using the campaign finance records from the 2003-2004 election cycle. Using the posterior mode to categorize the ideological affiliations of the political committees, our estimates match the self-reported ideology for 94.63% of those committees who self-reported to be Democratic and 89.49% of those committees who self-reported to be Republican.

Degree Type

Dissertation

Degree Name

Doctor of Philosophy (PhD)

Graduate Group

Economics

First Advisor

Hanming Fang

Subject Categories

Economics

ESSAYS ON THE ECONOMICS OF NETWORKS

Yiran Chen

A DISSERTATION

in

Economics

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2019

Supervisor of Dissertation

____

Hanming Fang
Professor of Economics

Graduate Group Chairperson

____

Jesús Fernández-Villaverde
Professor of Economics

Dissertation Committee

Hanming Fang, Professor of Economics

Andrew Postlewaite, Professor of Economics and Professor of Finance

Petra Todd, Professor of Economics

ESSAYS ON THE ECONOMICS OF NETWORKS

Dedicated to my parents.

ACKNOWLEDGMENT

I am deeply indebted to Professors Hanming Fang, Andrew Postlewaite, and Petra Todd for their guidance, support, and encouragement throughout my time at Penn. In particular, I would like to extend my gratitude to my advisor, Professor Fang, for having a profound belief in my work and for encouraging me to take on challenges. The completion of my dissertation would not have been possible without his support and nurturing. I am also grateful to other Professors who have helped me in various stages of my dissertation, including Kenneth Burdett, Francis J. DiTraglia, Andrew Shephard, Xun Tang, and Rakesh Vohra.

I would like to extend my sincere thanks to my friends at Penn including Xiang Fang, Qing Gong, Daniel Hauser, Yunan Li, Hongxun Ruan, Marjorie Rutherford, Paul Sangrey, Takeaki Sunada, Yu Wang, Yiwen Wu, YinYin Yu, and Weilong Zhang for helpful discussions as well as moral support. Thanks also to my long-time friends, especially Shan, Jiajun, Yi, Lingwei, and You, for their enduring friendship.

Finally, I would like to extend my special thanks to my parents for their continuous and unparalleled love. I am indebted to them for providing me with opportunities and experiences that have made me who I am and for selflessly encouraging me to pursue my dreams. I wouldn't have embarked on this journey without their support, and I dedicate this dissertation to them.

ABSTRACT

ESSAYS ON THE ECONOMICS OF NETWORKS

Yiran Chen

Hanming Fang

Empirical studies of social and economic networks are facilitated by the growing availability of network data. This area of research focuses on understanding two major questions: how networks affect economic outcomes and how networks are formed. This dissertation studies these questions respectively. The first chapter examines the impact of social networks on agents' economic outcomes in the context of job referrals in the labor market. The second chapter relates to the formation of financial networks with latent traits in the context of U.S. campaign contributions.

In the first chapter, "A Structural Analysis of Job Referrals and Social Networks: The Case of the Corporate Executives Market", I develop and structurally implement a labor market search model in which workers, in addition to directly receiving job offers, also receive referrals from their social contacts. In the model, referrals are generated endogenously: an external referral occurs when a friend rejects an offer he/she receives, and an internal referral occurs when a friend leaves his/her current job. I estimate the model by Generalized Method of Moments using data on the labor market history and the social connections of executives in S&P 500 firms. Using the estimated model, I find that referrals play a substantial role in the executive labor market. More than one quarter of the job transitions and raises are driven by

referrals. Shutting down referrals reduces executives' welfare by an equivalence of a two to seven percentage points reduction in income. I also evaluate the impacts of the social networks' structure by comparing the outcomes under the observed networks and alternative randomly formed networks. I find that the welfare distribution is more unequal under the random networks. I further investigate the mechanisms for these effects through the lens of two network statistics: friends' popularity and local community clustering.

In the second chapter, "Inferring the Ideological Affiliations of Political Committees via Financial Contributions Networks" (co-authored with Hanming Fang), we address the missing data problem for about two thirds of the political committees that do not self-identify their party affiliations in their registration with the Federal Election Commission. In this chapter, we propose and implement a novel Bayesian approach to infer the ideological affiliations of political committees based on the network of financial contributions among them. In Monte Carlo simulations, we demonstrate that our estimation algorithm achieves very high accuracy in recovering these committees' latent ideological affiliations when the pairwise difference in ideology groups' connection patterns satisfy a condition known as the Chernoff-Hellinger divergence criterion. We illustrate our approach using the campaign finance records from the 2003-2004 election cycle. Using the posterior mode to categorize the ideological affiliations of the political committees, our estimates match the self-reported ideology for 94.63% of those committees who self-reported to be Democratic and 89.49% of those committees who self-reported to be Republican.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

Introduction

Empirical studies of social and economic networks are facilitated by the growing availability of network data. This area of research focuses on understanding two major questions: how networks are formed, and how networks affect economic outcomes. This dissertation consists of two chapters and studies these questions respectively. The first chapter examines the impact of social networks on agents' economic outcomes in the context of job referrals in the labor market. This study focuses on workers' decisions given the observed networks and evaluates the welfare implications of different network configurations. The second chapter relates to the formation of financial networks with latent traits in the context of U.S. campaign contributions. This study uses a network formation model and observed network to recover the latent traits of the economic agents.

In the first chapter, "A Structural Analysis of Job Referrals and Social Networks: The Case of the Corporate Executives Market", I analyze quantitatively the value of social networks in the context of job referrals in the labor market. The existing literature has documented the importance of job referrals and social connections, but most of them only focus on the effect of direct connections, i.e., who you know. The main contribution of this chapter is evaluating the impact of the full social network structure.

Network structure beyond direct connections provides useful and interesting insights. For example, the value of a friend may depend on the number of friends he/she has. Additionally, it is not clear whether a friend with many connections is more valuable than a friend with few connections because on the one hand, a friend with many connections is in a good position to provide help, on the other hand, however, it also leads to stronger competition for his/her attention and help. Theoretically, the relative importance of these two forces is ambiguous, and thus it requires an empirical evaluation. This is an example in which indirect connections have a non-trivial effect on the value of direct connections. To study questions like this, I develop and implement a general framework to quantify the impact of the full social network structure in the labor market.

I build a labor search model with job referrals, which generates the two forces mentioned before as well as other network effects. The empirical environment is the senior executive labor market in the U.S., specifically, the C-suite in S&P 500 companies. This is a market where referrals could potentially play an important role because typically people cannot directly apply for these positions. Referrals provide candidates with additional chances to meet with firms. I combine data from three sources, including information on executives' labor market history and their social networks, and estimate the model using Generalized Method of Moments.

The key findings are the following. First, I find referrals to be important both statistically and economically. A model specification test rejects a model without referral, and a simulation shows that more than a quarter of job transitions and wage raises are driven by referrals. Second, I evaluate the impact of social network structure. At an individual level, I provide quantitative answers to the following questions: (1) whether a well-connected friend is beneficial and (2) whether connections among a person's friends are beneficial. Additionally, at a global level, I investigate how the change of network structure affects the distribution of executives' welfare and further investigate the mechanism.

In the second chapter, "Inferring the Ideological Affiliations of Political Committees via Financial Contributions Networks" (co-authored with Hanming Fang), we develop and apply a method to recover the latent traits of economic agents through network data and a formation model in the context of U.S. campaign finance.

In campaign finance, a large number of organizations send and receive political donations. Theses organizations include election campaigns, political action committees, and lobbyists, among others, which we will refer to as political committees. About two thirds of these committees do not self-identify their party affiliation in their registration with the Federal Election Commission, and the goal of this paper is to recover their latent political ideology. It is important to understand these political committees' ideologies because they are the direct collectors of individual political donations. In order to understand the behavior of individual donors, it is necessary to understand the organizations to which they contribute.

To this end, we build a network formation model based on the Stochastic Block Model and estimate the model by Markov Chain Monte Carlo (MCMC) method using the observed political donation networks.

In Monte Carlo simulations, we demonstrate that our estimation algorithm achieves very high accuracy in recovering these organizations' latent ideological affiliations when the pairwise difference in ideology groups' connection patterns satisfy a condition known as the Chernoff-Hellinger divergence criterion. We illustrate our approach using the campaign finance records from the 2003-2004 election cycle. Using the posterior mode to categorize the ideological affiliations of the political committees, our estimates match the self-reported ideology for 94.63% of those committees who self-reported to be Democratic and 89.49% of those committees who self-reported to be Republican.

Our method contributes to the political economy literature in the following aspects. First, it does not assume a bipartite graph with mutually exclusive sets of donors and recipients. Second, it is valid with one observation of the network. Third, it does assume homophily. In other words, it does not require political committees to be most likely to donate to other committees with the same ideology, which is potentially a problematic assumption for the politically independent. Additionally, our method contributes to the community detection literature in the following aspects. First, it accommodates rich vertex-level and edge-level observable characteristics, and it incorporates the strength of the edge. Second, it incorporates the self-reported political affiliations from a subset of the committees.

Chapter 1

A Structural Analysis of Job Referrals and Social Networks: The Case of the Corporate Executives Market

1.1 Introduction

Job referrals are an important channel through which workers find jobs and firms fill vacancies.[1] Because of referrals, workers who are more socially interconnected may experience better labor market outcomes and advance more quickly in their careers. In this paper, I study the network effects of job referrals and investigate the impacts of social network structure on workers' labor market outcomes. I develop a job search model that incorporates referral and non-referral job offers and different kinds of social networks. The network structure is dynamic and evolves as workers move across jobs. To estimate the model, I combine three different data sources on the corporate executive labor market.

In my model, workers directly receive job offers as well as referrals from their friends. Referrals are generated endogenously: an external referral occurs when a friend rejects an offer that he/she receives, and an internal referral occurs when a friend leaves his/her current job. As a result of this referral process, the quantity and

[1] For example, Granovetter (1973) reports that 56% of the job seekers in a Boston suburb obtained their jobs through social contacts. Other surveys at different locations and times report numbers ranging from 25% to 87%. Marsden and Gorman (2001) report that 37% of U.S. firms often use referral in recruitment. Other surveys at different locations and times report numbers ranging from 36% to 88%. See Topa (2011) for a detailed summary.

the quality of referrals depend not only on the worker's number of friends but also on the quality of these friends' jobs. Moreover, the model generates rich network effects beyond immediate friends. First, the *popularity of a friend* affects referrals. On the one hand, a popular friend means high competition for referrals, lowering a worker's probability of receiving a particular referral sent by his/her popular friend (*competition effect*). On the other hand, a popular friend benefits from his/her large set of friends' referrals, increasing the quantity and the quality of referrals he/she sends out (*ripple effect*). Second, *local clustering*, defined as the fraction of a worker's friends who are also friends with one another, also affects referrals. An advantage of high clustering is that it keeps the positive spillovers in an inner circle (*closeness effect*). A disadvantage is that it limits the positive spillovers from a distance (*isolation effect*).

My empirical analysis is based on three data sets: (1) Compustat Executive Compensation (ExecuComp), (2) BoardEx, and (3) U.S. Stock Database from the Center for Research in Security Prices (CRSP). The first data set is used to construct a panel of individuals' employment history in the executive market along with their on-the-job compensation. The second data set is used to construct three social networks representing social connections established via education, work, and other social activities. The third data set provides firms' financial variables.

My analysis first provides reduced-form evidence on job referrals in the executive labor market. First, I show that socially connected executives' compensation is positively correlated, controlling for time-invariant individual characteristics, time-specific shocks, and industry-specific shocks. This result supports my model's prediction that executives with better jobs send better quality referrals, increasing their friends' compensation. Second, I show that an individual is more likely to make a career advancement when his/her executive friends leave their current jobs, controlling for time-invariant individual characteristics and time-specific shocks. This finding suggests that executives who leave their current jobs send referrals that help their friends advance in their careers.

I then estimate the structural job search model by Generalized Method of Moments (GMM). The estimation results show both the statistical and economical significance of job referrals. My model nests a model with no referrals, and a specification test rejects such a model. Simulations from my model provide a way of assessing the importance of referrals in the job market dynamics in this labor market. A decomposition shows that 27.86% of non-executive to executive transitions, 66% of executive job-to-job transitions, and 82.1% of raises are driven by referrals.

I use the estimated model to perform two counterfactual experiments. The first experiment evaluates the welfare effect of referrals by varying the probability of referrals. I find that shutting down referrals reduces executives' welfare by an equivalence of a 2-7% decrease in annual income and that increasing referral probability to one boosts executives' welfare by an equivalence of a 6-16% increase in annual income.

The second experiment examines the welfare effect of network structure by varying the network structure. Specifically, a new set of counterfactual networks are generated in which the individuals have the same number of friends as the observed networks, but the connections are formed randomly. I find that the welfare distribution is more unequal under the randomly formed networks. I further investigate the mechanisms for these effects through the lens of two network statistics: friends' popularity and local clustering. First, in terms of friends' popularity, the *competition effect* dominates the *ripple effect*. The random networks increase friends' popularity for individuals with a small number of friends and decrease friends' popularity for individuals with a large number of friends. Therefore, the increased competition for referral makes individuals with a small number of friends worse off, and the decreased competition for referral makes individuals with a large number of friends better off, which increases inequality. Second, in terms of local clustering, the *isolation effect* dominates for individuals with a small number of friends, and the *closeness effect* dominates for individuals with a large number of friends. The random networks universally decrease the local clustering. Therefore, the decreased isolation effect makes individuals with a small number of friends better off, and the decreased closeness effect makes individuals with a large number of friends worse off, which decreases inequality. Overall, the competition effect resulting from the change in friend popularity dominates, generating greater inequality under the random networks. This experiment highlights the effects of network structures beyond the number of friends.

My paper contributes to three strands of the literature. First, it contributes to studies of the impact of social connections on executive compensation (e.g., Shue, 2013; Engelberg et al., 2013) by building a model that formalizes the mechanisms by which social connections impact executives' compensation and job transitions. Second, it contributes to studies of labor search models with job referrals (e.g., Montgomery, 1992; Mortensen and Vishwanath, 1994; Calvó-Armengol and Zenou, 2005; Ioannides and Soetevent, 2006; Galenianos, 2014; Arbex et al., 2018) by incorporating the full structure of the network (beyond the number of friends). The model presents a general framework to study a rich set of network effects of job referrals. Third, it contributes to studies of labor market dynamics on social networks (e.g., Topa, 2001; Calvó-Armengol and Jackson, 2004; Calvó-Armengol and Jackson, 2007) by modeling micro-founded workers' decisions and wage bargaining processes.

Literature Review

My paper is related to four broad strands of the literature. First, it complements the literature on labor market peer effect and job referrals in the general labor market.[2] Surveys show that referrals are frequently used by workers to find jobs (e.g., Granovetter, 1973; Pellizzari, 2010) and by firms to fill vacancies (e.g., Marsden and Gorman, 2001).[3] Additionally, empirical studies (e.g., Topa, 2001; Conley and Topa, 2007; Schmutte, 2014) show that local job referrals generate positive spatial correlation in workers' employment statuses and wage premia. More recently, Burks et al. (2015) and Brown et al. (2016) characterize the relationships among referrals, match quality, wage trajectories, and turnover using data sets with direct information on referrals. In addition to documenting empirical patterns, some studies develop theoretical models on referrals, starting with the pioneering work by Boorman (1975). Some studies model referrals as means to reduce search friction by providing more opportunities for workers and firms to meet. Others model referrals as means to reduce information friction by providing firms with more information on worker quality or match quality (e.g., Rees, 1966; Montgomery, 1991; Arrow and Borzekowski, 2004; Dustmann et al., 2015). My paper focuses on the role referral plays in mitigating search friction, and I discuss related papers in reviewing the search literature.

Second, my paper contributes to the literature on executive compensation.[4] In

[2] See Ioannides and Datcher Loury (2004) and Topa (2011) for comprehensive surveys.

[3] See Holzer (1988), Corcoran et al. (1980), Granovetter (1995), Gregg and Wadsworth (1996), and Addison and Portugal (2002) for additional evidence on workers' usage of referrals; see Holzer (1987), Neckerman and Kirschenman (1991), Miller and Rosenbaum (1997), Fernandez et al. (2000), and Brown et al. (2016) for additional evidence on firms' usage of referrals. A more comprehensive and detailed listing can be found in Topa (2011).

[4] See Frydman and Jenter (2010) and Edmans and Gabaix (2016) for comprehensive surveys.

the literature, some researchers view the level of total compensation as a competitive outcome (e.g., Gabaix and Landier, 2008; Terviö, 2008), while others view it as rent extraction and managerial entrenchment (e.g., Bebchuk and Fried, 2004; Kuhnen and Zwiebel, 2008). My model combines both views by incorporating competition from outside offers and allowing executives to earn positive rent through bargaining. Because the focus of my study is job search and referrals, my paper abstracts from problems arising from asymmetric information such as learning about executive ability (e.g., Taylor, 2013) or contracting in the presence of hidden action and hidden information (e.g., Gayle et al., 2015; Gayle et al., 2016).

Among the literature on executive compensation, my paper directly contributes to studies examining the impact of executives' social connections on their compensation. Previous studies focus mostly on documenting the empirical relationship between social connection and compensation.[5] For example, Shue (2013) uses the random assignment of MBA students to sections at the Harvard Business School to show that executive compensation is significantly more similar among graduates from the same section than among graduates from different sections, and that this effect is more than twice as strong in the year following alumni reunions. Engelberg et al. (2013) investigate CEOs' social connections to other executives and directors outside their firms and show that these connections increase CEO compensation, and that the increase is higher for connections with "important" people such as CEOs of big firms. My paper contributes to this literature by providing additional empirical evidence, and moreover, a structural model that formalizes the mechanisms by which social connections impact executives' compensation and job transitions.

[5]There is another set of studies examining the impact of executives and directors' social connections on firm performance. See El-Khatib et al. (2015), Hwang and Kim (2009), Fracassi and Tate (2012), Cai and Sevilir (2012), Larcker et al. (2013), and Ruan (2017).

Third, my paper contributes to the theoretical and empirical literature on labor search models.[6] There is a large body of literature on sequential random job search. For example, the wage posting model in Burdett and Mortensen (1998) is adapted and empirically implemented in Van den Berg and Ridder (1998), Bontemps et al. (1999), Bontemps et al. (2000), Meghir et al. (2015), Shephard (2017), and Aizawa and Fang (2018). The wage bargaining model in Postel-Vinay and Robin (2002) and Cahuc et al. (2006) is adapted and empirically implemented in Dey and Flinn (2005) and Bagger et al. (2014). My model is based on that of Cahuc et al. (2006), in which workers search both on and off the job, climb the productivity ladder, and bargain their wages with the firms.

Among the literature on labor search, my paper directly contributes to studies of referrals in search-based frameworks, in which referral is modeled as an additional channel of job arrival. Previous studies do not accommodate rich social network structure and generate very limited network effects. In early studies such as Montgomery (1992) and Mortensen and Vishwanath (1994), there is no notion of social connections. In later studies such as Calvó-Armengol and Zenou (2005), Ioannides and Soetevent (2006), Galenianos (2014), and Arbex et al. (2018), referrals are sent through social networks. However, they do not incorporate network structure beyond the degree distribution (i.e., the number of social connections), and they impose strong assumptions on the network structure to achieve this tractability.[7] My paper

[6]See Rogerson et al. (2005) for a survey of theoretical studies and Eckstein and Van den Berg (2007) for a survey of empirical studies.

[7]In Calvó-Armengol and Zenou (2005) and Ioannides and Soetevent (2006), a worker randomly draws new friends in every period, so only the immediate friends matter. In Galenianos (2014), a worker has a continuum of friends, so the unemployment rate of one's friends does not change in the steady state. In Arbex et al. (2018), the random network formation model, combined with the worker's information structure, implies that a worker's degree is a sufficient statistic for his/her future value.

differs from this set of papers in that it does not impose assumptions on how the networks are formed, and that it uses the full structure of the network (the adjacency matrices), which generates rich network effects. I highlight the importance of the network structure beyond the number of friends by showing in my counterfactual experiment that networks with the same degree sequence but different connection patterns lead to different welfare distribution. Additionally, my model allows for a rich characterization of referrals, both internal and external. In Calvó-Armengol and Zenou (2005) and Ioannides and Soetevent (2006), there are only external referrals. In their models, jobs are identical, so employed workers send referrals whenever they receive outside job offers, while in my paper jobs are heterogeneous, and employed workers only pass along unwanted jobs. In Galenianos (2014) and Arbex et al. (2018), there are only internal referrals. In their models, employed workers randomly send referrals identical to their own jobs, while in my paper this only occurs when a worker leaves his/her current job, a modeling choice driven by the small number of executive positions.

Finally, my paper contributes to studies on the labor market dynamics on social networks such as Topa (2001), Calvó-Armengol and Jackson (2004), Calvó-Armengol and Jackson (2007), and Fontaine (2008).[8] These studies model job transitions of socially connected workers as a Markov process on a network, with the full network structure incorporated. Topa (2001) and Calvó-Armengol and Jackson (2004) focus on employment dynamics. In their models, an employed worker randomly contacts an unemployed friend and then the friend becomes employed. Calvó-Armengol and Jackson (2007) further incorporate wage dynamics. These three papers take a statistical approach and abstract from workers' decision-making problem and firms'

[8]See Jackson et al. (2017) for a survey of the broad literature on social networks.

wage-setting problem. My paper complements their studies by formally formulating workers' decision problem and the wage bargaining process, providing micro-founded employment and wage dynamics. In Fontaine (2008), wage is bargained, but the network is assumed to be complete, while my model accommodates any network configuration.[9] To the best of my knowledge, there are not many empirical studies of job referrals with explicit network models.[10] My paper contributes to this literature by developing a tractable yet flexible model with micro foundation for empirical analysis.

The remainder of the paper is structured as follows. Section 1.2 presents a job search model with referrals. Section 1.3 describes the data and shows reduced-form evidence. Section 1.4 describes the estimation strategy. Section 1.5 presents the estimation results. Section 1.6 discusses the counterfactual experiments. Section 1.7 concludes.

1.2 Model

Social Networks

In this section I describe the social networks through which referrals are sent. There are n workers in the labor market, indexed by i. Among them, there are three types of social connections established via overlapping experience in school, work, and social activities. Connections are undirected, and new ones are formed over time. These three types of connections constitute three networks. In these networks, a node represents a worker and an edge represents a social connection. At time t, the

[9] In a complete network, every pair of distinct nodes is connected. A complete network simplifies analysis because it eliminates heterogeneity in network position.

[10] Topa (2001) is one such empirical paper. As discussed, it does not feature micro-founded worker's decision problem and wage-setting process.

networks are characterized by adjacency matrices

$$Y^{Edu,t} = \{y_{ij}^{Edu,t}\}_{1\leq i,j\leq n},\ Y^{Work,t} = \{y_{ij}^{Work,t}\}_{1\leq i,j\leq n},\ Y^{Social,t} = \{y_{ij}^{Social,t}\}_{1\leq i,j\leq n},$$

where $y_{ij}^{Edu,t} = 1$ if workers i and j have established a school connection by time t, and $y_{ij}^{Edu,t} = 0$ otherwise; the same holds true for $y_{ij}^{Work,t}$ and $y_{ij}^{Social,t}$.

Both $Y^{Edu,t}$ and $Y^{Social,t}$ are exogenous. Part of $Y^{Work,t}$ is exogenous, and part of $Y^{Work,t}$ is endogenous as a result of transition in the executive labor market.[11] I assume that when individuals make labor market decisions, they do not consider the implied changes to the endogenous work network. This assumption is an implication of Assumption 1 (static expectation) which will be discussed later.

Additionally, define a simplified network that does not distinguish the different types of connections

$$Y^{All,t} = \{y_{ij}^{All,t}\}_{1\leq i,j\leq n},$$

where $y_{ij}^{All,t} = \max\{y_{ij}^{Edu,t}, y_{ij}^{Work,t}, y_{ij}^{Social,t}\}$, i.e., $y_{ij}^{All,t} = 1$ as long as there is some kind of social connection between i and j.

For any network $k \in \{Edu,\ Work,\ Social,\ All\}$, define worker i's friends in network k at time t as the set of nodes it is directly connected to:

$$N^{k,t}(i) = \{j \mid y_{ij}^{k,t} = 1\}. \tag{1.1}$$

Define worker i's degree in network k at time t as the number of i's friends:

$$d_i^{k,t} = |N^{k,t}(i)|. \tag{1.2}$$

[11]Endogenous connections are established between the executives in the same company. Other workplace connections are viewed as exogenous if one of the individuals is not an executive. For example, when executives in different companies serve on the same board of directors in a third company, this connection is viewed as exogenous to the model of executive labor market.

Finally, define worker i's local clustering coefficient in network k at time t as the number of connections between i's friends divided by the number of possible connections between them:

$$c_i^{k,t} = \begin{cases} \frac{|\{y_{jl}^{k,t}: j,l \in N^{k,t}(i), y_{jl}^{k,t}=1\}|}{\binom{d_i^{k,t}}{2}} & \text{if } d_i^{k,t} > 1 \\ 0 & \text{otherwise} \end{cases}, \tag{1.3}$$

where $d_i^{k,t}$ and $N^{k,t}(i)$ are defined in (1.2) and (1.1). Local clustering coefficient is a measure of how tightly connected a local network is. It takes a value between $[0, 1]$, characterizing the fraction of worker i's friends who are also friends with one another.

To avoid verbosity, throughout the paper, when I refer to friend, degree, or local clustering coefficient without specifying a particular network, I mean those associated with the simplified network Y^{All}. Additionally, to ease notation, I omit time superscript t when it does not cause confusion.

Job Search with Referrals

In this section I describe the job search model with referrals through social connections. I follow the terminology in the standard search literature, but certain terms should be interpreted in the context of the executive labor market: specifically, "workers" refer to executives and candidates for executive jobs, "wage" refers to the value of the compensation package, and "unemployment" refers to the status of not working in an executive job. The basic settings are similar to Cahuc et al. (2006). A novel feature in my model is that workers can send referrals to friends after they reject an offer or leave their current jobs, which affects both job transition and wage bargaining.

I consider a labor market with a finite number of workers and firms. Workers and firms are matched randomly in a frictional labor market. Time is continuous. Workers enjoy instantaneous utility $U(x)$ from income x and discount the future at rate ρ.

Production. Workers differ in their abilities (denoted as a) as well as their positions in the social networks. Firms differ in their productivities (denoted as p). A firm is modeled as a collection of jobs with the same productivity. The marginal product for a worker-firm pair (a, p) is ap. An unemployed worker receives an income flow of ab, which he/she has to forgo upon finding a job.

Direct Job Arrivals and Job Separations. Firms and workers meet randomly. Unemployed workers receive direct job offers at Poisson rate λ_0, and employed workers at rate λ_1. The productivity p of the firm from which an offer originates is randomly distributed on $[p^{min}, p^{max}]$ according to cdf $F(p)$. Exogenous job separations occur at Poisson rate δ.

Referrals. Referrals are by-products of traditional types of labor market transitions. They always follow direct job arrivals or job separations, when workers reject offers or leave their current firms.[12] Specifically, there are three situations under which a worker sends a referral:

1. External referral following direct job arrival: if a worker rejects a direct offer, he/she may refer one of his/her friends to the firm he/she declines. I call it an external referral because it is initiated by an individual outside of the firm.

[12] In this model, a worker cannot refer a friend to his/her current firm unless he/she himself/herself leaves. This modeling choice is driven by the small number of executive positions in each firm.

2. Internal referral following direct job arrival: if a worker leaves his/her current firm because he/she accepts a better job offer, he/she may refer one of his/her friends to the firm he/she leaves. I call it an internal referral because it is initiated by an individual who has worked at the firm.

3. Internal referral following job separation: if a worker leaves his/her firm as a result of exogenous separation, he/she may refer one of his/her friends to the firm he/she leaves.[13] This is also an internal referral.

Workers are assumed to be nonstrategic in sending referrals. They send referrals with probability π_1 after direct job arrivals and with probability π_0 after job separations. Conditional on sending a referral, a recipient is sampled according to the following sequential statistical process:

1. Sample employment status: sample unemployed friends with probability ν_u and employed friends with probability $\nu_e = 1 - \nu_u$.

2. Sample a network: sample school friends, work friends, and social-activity friends with probability $(\omega^{Edu}, \omega^{Work}, \omega^{Social}) \in \Delta^2$. When not all three types of friends are present, only sample from the available types and normalize the probability accordingly.

3. Sample a friend: conditional on employment status and type of network, sample one friend randomly.

[13] Job separations can be either involuntary or voluntary. Voluntary separations such as retirement or leaving for health reasons may lead to internal referrals.

More precisely, conditional on friend $j \in N^{All}(i)$ sending a referral, the probability for worker i to receive this referral is

$$\gamma_{i \leftarrow j} = \begin{cases} \nu_u(\xi_j^{Edu \cap u} \cdot \frac{y_{ij}^{Edu}}{d_j^{Edu \cap u}} + \xi_j^{Work \cap u} \cdot \frac{y_{ij}^{Work}}{d_j^{Work \cap u}} + \xi_j^{Social \cap u} \cdot \frac{y_{ij}^{Social}}{d_j^{Social \cap u}}) & \text{if } s_i = u, \\ \nu_e(\xi_j^{Edu \cap e} \cdot \frac{y_{ij}^{Edu}}{d_j^{Edu \cap e}} + \xi_j^{Work \cap e} \cdot \frac{y_{ij}^{Work}}{d_j^{Work \cap e}} + \xi_j^{Social \cap e} \cdot \frac{y_{ij}^{Social}}{d_j^{Social \cap e}}) & \text{if } s_i = e, \end{cases} \tag{1.4}$$

where $\mathbf{s} = \{s_i\}_{i=1,\ldots,n}$ is a vector of employment statuses for all workers. For each network $k \in \{Edu, Work, Social\}$. $d_j^{k \cap u} = |N^k(j) \cap \{i : s_i = u\}|$ is the number of j's unemployed friends in network k, and $d_j^{k \cap e} = |N^k(j) \cap \{i : s_i = e\}|$ is the number of j's employed friends in network k. $\xi_j^{k \cap u} = \frac{\mathbb{1}(d_j^{k \cap u} > 0)\omega^k}{\mathbb{1}(d_j^{Edu \cap u} > 0)\omega^{Edu} + \mathbb{1}(d_j^{Work \cap u} > 0)\omega^{Work} + \mathbb{1}(d_j^{Social \cap u} > 0)\omega^{Social}}$ is the probability for j to send referrals to unemployed friends in network k, and $\xi_j^{k \cap e} = \frac{\mathbb{1}(d_j^{k \cap e} > 0)\omega^k}{\mathbb{1}(d_j^{Edu \cap e} > 0)\omega^{Edu} + \mathbb{1}(d_j^{Work \cap e} > 0)\omega^{Work} + \mathbb{1}(d_j^{Social \cap e} > 0)\omega^{Social}}$ is the probability for j to send referrals to employed friends in network k.

In this model, a referral is a chance for a worker to meet with a firm: an additional source of job arrival. It is tied to the firm's productivity, not to the wage offered to the worker sending the referral. All the wages are bargained between worker-firm pairs.

I make the following additional assumptions of the referral process. First, sending a referral is costless and takes no time. Second, unemployed workers do not send referrals when they reject job offers.[14] Third, referrals have no immediate "chain effect": when a worker rejects a referral, he/she no longer passes it along to his/her friends; when a worker accepts a referral and change jobs, he/she does not send an internal referral about the job he/she leaves.[15] This assumption is supported by reduced-form

[14]This assumption guarantees the tractability of the model. It ensures that the "reservation productivity" for an unemployed worker is a single agent decision problem. Otherwise, if unemployed workers are allowed to send referrals after rejecting job offers, the "reservation productivities" for unemployed workers will be interdependent, resulting in a game among the workers. In terms of the empirics, the estimation results show few unemployed workers reject offers. Therefore, this assumption is relatively innocuous quantitatively.

[15]This assumption reduces the computational intensity of the model. Relaxing this assumption

evidence in Section 1.3 that there is no significant co-movement in job transitions beyond immediate friends.

Wage Bargaining. Wages are bargained over by workers and firms, and the bargaining process is the same for direct offers and referrals. In the bargaining, information is complete. All participating agents observe one another's types: the firm's productivity p, and the worker's ability a as well as his/her state variable Γ which embodies information relevant to the arrivals of referrals. For expository purposes, I focus on bargaining here and defer the description of Γ to later. Firms can vary their wage offers according to the characteristics (a, Γ) of the particular worker they meet. They can also counter the offers received by their employees from competing firms. Additionally, wage contracts are long-term contracts that can be renegotiated by mutual agreement only. Finally, the bargaining outcome is such that a worker obtains his/her reservation value and a share β of the additional worker rent, where β represents the worker's bargaining power.[16]

Formally, let $V_0(a, \Gamma)$ denote the lifetime value of an unemployed worker with ability a and state variable Γ; let $V_1(a, w, p, \Gamma)$ denote that of the same worker when employed at a firm of productivity p and paid a wage w. The bargained wage between a type-(a, Γ) unemployed worker and a type-p firm, denoted as $\phi_0(a, p, \Gamma)$, satisfies

$$V_1(a, \phi_0(a, p, \Gamma), p, \Gamma) = V_0(a, \Gamma) + \beta[V_1(a, ap, p, \Gamma) - V_0(a, \Gamma)]. \tag{1.5}$$

will not substantially change the analytical property of the model. It is useful to note that this assumption is not as restrictive as it seems. Although in the short run referrals only affect immediate friends, in the long run the network effects will propagate beyond immediate friends (the ripple effect).

[16] When the worker's utility is linear, this can be interpreted as a Nash bargaining solution. In more general cases, Cahuc et al. (2006) show that this is the outcome of a strategic bargaining game adapted from Rubinstein (1982).

In this equation, $V_0(a,\Gamma)$ is the worker's reservation value, $V_1(a,ap,p,\Gamma)$ is the maximum value the worker can hope to extract from the match with the firm, and $[V_1(a,ap,p,\Gamma)-V_0(a,\Gamma)]$ is the additional worker rent brought about by this match. Note that the worker only accepts the offer if $V_1(a,ap,p,\Gamma)\geq V_0(a,\Gamma)$. Equivalently, this can be characterized by a productivity threshold $p_0(a,\Gamma)$, defined by

$$V_1(a,ap_0(a,\Gamma),p_0(a,\Gamma),\Gamma)=V_0(a,\Gamma). \tag{1.6}$$

If $p\geq p_0(a,\Gamma)$, the worker accepts the offer; otherwise, he/she rejects it.

When a worker employed at firm p is contacted by an outside firm p', the incumbent firm and the poaching firm compete for the worker. If the poaching firm is more productive than the incumbent ($p'>p$), it wins the bargain by offering a wage $\phi_1(a,p,p',\Gamma)$ such that

$$V_1(a,\phi_1(a,p,p',\Gamma),p',\Gamma)=V_1(a,ap,p,\Gamma)+\beta[V_1(a,ap',p',\Gamma)-V_1(a,ap,p,\Gamma)]. \tag{1.7}$$

In this case, the worker's reservation value is the maximum value he/she can extract from the incumbent firm, $V_1(a,ap,p,\Gamma)$.

If the poaching firm is less productive than the incumbent ($p'\leq p$), the incumbent retains the worker by offering a renegotiated wage $\phi_1(a,p',p,\Gamma)$ such that

$$V_1(a,\phi_1(a,p',p,\Gamma),p,\Gamma)=V_1(a,ap',p',\Gamma)+\beta[V_1(a,ap,p,\Gamma)-V_1(a,ap',p',\Gamma)]. \tag{1.8}$$

Note that renegotiation requires mutual agreement, so it is only triggered if $\phi_1(a,p',p,\Gamma)$ is higher than the worker's current wage w. Equivalently, this can be characterized by a productivity threshold $q(a,w,p,\Gamma)$, defined by

$$\phi_1(a,q(a,w,p,\Gamma),p,\Gamma)=w. \tag{1.9}$$

If the poaching firm's productivity is relatively high ($q(a, w, p, \Gamma) < p' \leq p$), the worker gets a raise in the incumbent firm; if the poaching firm's productivity is too low ($p' \leq q(a, w, p, \Gamma)$), the outside offer is discarded.

Labor Market Transitions and Referrals. With the descriptions above, I can fully characterize the worker's labor market transitions and referrals.

1. When an unemployed type-(a, Γ) worker meets a firm p, either from direct arrival or referral,

   a) If $p \geq p_0(a, \Gamma)$, the worker accepts the offer at a bargained wage $\phi_0(a, p, \Gamma)$;

   b) Otherwise, the worker rejects the offer.

2. When an employed type-(a, Γ) worker at firm p paid at wage w meets an outside firm p', either from direct arrival or referral,

   a) If $p' > p$, the worker moves to the new firm with a wage $\phi_1(a, p, p', \Gamma)$. Additionally, if this follows a direct job arrival, the worker may refer a friend to the firm he/she leaves (internal referral).

   b) If $q(a, w, p, \Gamma) < p' \leq p$, the worker stays at the current firm with a raise to $\phi_1(a, p', p, \Gamma)$.
      Additionally, if this follows a direct job arrival, the worker may refer a friend to the firm he/she rejects (external referral).

   c) If $p' \leq q(a, w, p, \Gamma)$, the worker keeps the current wage at the current firm. Additionally, if this follows a direct job arrival, the worker may refer a friend to the firm he/she rejects (external referral).

3. When an employed worker is exogenously separated from his/her current job, he/she may refer a friend to the firm he/she leaves (internal referral).

State Variable Γ^i. $\Gamma^i = \{\Gamma_0^i(\cdot), \Gamma_1^i(\cdot)\}$ are distributions of the productivities of worker i's friends' jobs, when i is unemployed and employed respectively.[17] More precisely, $\Gamma_0^i(p)$ and $\Gamma_1^i(p)$ are the "head count" of i's friends working at firms with productivities no greater than p, where each friend j is weighted by $\gamma_{i\leftarrow j}$, his/her probability of sending referral to i. As $\gamma_{i\leftarrow j}$, defined in (1.4), depends on recipient i's employment status, so does Γ^i. Specifically, Γ_0^i is calculated using $\gamma_{i\leftarrow j}$ for unemployed i, and Γ_1^i is calculated using $\gamma_{i\leftarrow j}$ for employed i:

$$\begin{aligned}\Gamma_0^i(p) &= \sum_{j\in N^{All}(i)} \gamma_{i\leftarrow j}(s_i = u)\cdot \mathbb{1}(p_j \le p),\\ \Gamma_1^i(p) &= \sum_{j\in N^{All}(i)} \gamma_{i\leftarrow j}(s_i = e)\cdot \mathbb{1}(p_j \le p).\end{aligned} \tag{1.10}$$

In the following part, I discuss two properties of Γ^i. First, $(\lambda_1, \pi_1, \delta, \pi_0, F, \Gamma^i)$ fully characterize worker i's *instantaneous* arrival rate and distribution of referrals. To illustrate this point, first consider the referrals from one particular friend j. External referral from j arrives at rate $\lambda_1\pi_1 F(p_j)\gamma_{i\leftarrow j}$, and the associated productivity distribution is F truncated above at p_j. Internal referral from j arrives at rate $[\lambda_1\pi_1(1-F(p_j)) + \delta\pi_0]\gamma_{i\leftarrow j}$, and the associated productivity is p_j. Then use Γ^i to aggregate over friends.

Second, $(\lambda_1, \pi_1, \delta, \pi_0, F, \Gamma^i)$ are not sufficient statistics to forecast the dynamics of Γ^i. The dynamics of Γ^i depend on the dynamics of friends' jobs $\{p_j\}_{j\in N^{All}(i)}$. Friends' jobs are affected by referrals they receive, i.e. $\{\Gamma^j\}_{j\in N^{All}(i)}$, and thus their friends jobs

[17] Rigorously speaking, they are not probability distributions because the total measures do not necessarily equal one.

$\{\{p'_j\}_{j' \in N^{All}(j)}\}_{j \in N^{All}(i)}$. By similar argument, the interdependence goes further on the network (ripple effects). Therefore, the dynamics of Γ^i is determined by the structure of the full networks and the dynamics of all workers' jobs.[18] Even though referrals do not have immediate effect beyond direct friends, in the long run, referrals generate ripple effects that spread across the whole network.

Worker's Information Set. First, a worker i observes the following information about his/her local networks:

$$N^{Edu}(i), N^{Work}(i), N^{Social}(i);$$
$$\{d_j^{Edu\cap u}, d_j^{Edu\cap e}, d_j^{Work\cap u}, d_j^{Work\cap e}, d_j^{Social\cap u}, d_j^{Social\cap e}\}_{j \in N^{All}(i)}.$$ [19]

He/she knows the identities of his/her friends and the types of their connections. Additionally, he/she knows his/her friends' degree in unemployed and employed school network, work network, and social-activity network, so he/she can calculate the level of "competition" for referrals. Second, he/she observes the productivities of his/her friends' jobs:

$$\{p_j\}_{j \in N^{All}(i)}.$$

Therefore, a worker has enough information to calculate the instantaneous Γ^i.

Workers' Forward-Looking Behavior. I make the following assumptions on workers' forward-looking behavior.

Assumption 1. Workers have static expectation of Γ.

[18] For networks with multiple components, rigorously speaking, the dynamics of Γ^i is determined by the structure of the connected component containing i and the dynamics of the workers in the component.

[19] It is reasonable to assume that a worker knows only his/her local network. Previous studies (e.g., Friedkin, 1983; Krackhardt, 1987; Krackhardt, 2014; and Banerjee et al., 2017) show that people have little knowledge of their social network structure beyond immediate friends.

Workers do not forecast the dynamics of Γ. When workers bargain with firms, they calculate Γ using the latest information and assume it does not change. In other words, workers ignore future changes of their networks as well as their friends' jobs, and thus ignore future changes of the arrival rate and distribution of referrals in their calculation of future values. This assumption can be interpreted as bounded rationality. Given workers' limited information, it is forbiddingly difficult to calculate Γ's law of motion for two main reasons. First, Γ is high dimensional; second, the calculation requires integration over the structures of the unobserved part of the networks and integration over non-friends' jobs.[20]

This assumption introduces discrepancy between the worker's problem of solving bargained wage and my (the researcher's) problem of studying the dynamics. It should be emphasized that this assumption only applies to the worker's problem and does not apply to my study of the dynamics. I incorporate the evolution of the networks and analyze the labor market transitions of all workers according to the model description in the previous parts.

This assumption has two implications. First, in making labor market decisions, a worker does not consider the implied change to the endogenous work network. For example, a worker will not accept an "undesirable" job for the sole purpose of becoming friends with another worker in the same firm. This implication is not unrealistic, especially in the executive labor market, because an individual can achieve similar purpose by actively building on his/her social-activity network, whose evolution is accommodated in my model. Second, in sending referrals, a worker does not consider

[20] An alternative modeling approach is to approximate Γ with a low dimensional object and approximate its law of motion by imposing further assumptions. This alternative approach accommodates dynamics at the expense of accuracy. In this paper, I choose to use the accurate characterization of Γ and forgo the dynamics.

the implied change to friends' jobs. This implication is innocuous, especially for referrals following direct job offers. To illustrate, consider a situation when worker i refers a job to friend j and j accepts. This will not increase helpful referrals from j to i because the best referral from j is as good as j's own job, which i has forgone.

Assumption 2. Workers are rationally forward looking in all other aspects.

Given the arrivals of referrals, workers have rational expectations on future job arrivals, job separations, and bargained wages.

Worker's Problem and Bargained Wage. An unemployed worker's value function is characterized by

$$\rho V_0(a,\Gamma) = U(ab) \tag{1.11a}$$

$$+ \lambda_0 \int_{p_0(a,\Gamma)}^{p^{max}} [V_1(a,\phi_0(a,x,\Gamma),x,\Gamma) - V_0(a,\Gamma)]dF(x) \tag{1.11b}$$

$$+ \delta\pi_0 \int_{p_0(a,\Gamma)}^{p^{max}} [V_1(a,\phi_0(a,y,\Gamma),y,\Gamma) - V_0(a,\Gamma)]d\Gamma_0(y) \tag{1.11c}$$

$$+ \lambda_1\pi_1 \int_{p_0(a,\Gamma)}^{p^{max}} \int_{p_0(a,\Gamma)}^{y} [V_1(a,\phi_0(a,x,\Gamma),x,\Gamma) - V_0(a,\Gamma)]dF(x)d\Gamma_0(y) \tag{1.11d}$$

$$+ \lambda_1\pi_1 \int_{p_0(a,\Gamma)}^{p^{max}} \int_{y}^{p^{max}} [V_1(a,\phi_0(a,y,\Gamma),y,\Gamma) - V_0(a,\Gamma)]dF(x)d\Gamma_0(y), \tag{1.11e}$$

where $\phi_0(a,p,\Gamma)$ is the bargained wage defined in (1.5), and $p_0(a,\Gamma)$ is the reservation productivity defined in (1.6). To understand expression (1.11), note that line (1.11a) represents the flow utility; line (1.11b) represents the expected value of receiving a direct job offer; line (1.11c) represents the expected value of receiving an internal referral following friends' job separations; line (1.11d) represents the expected value of receiving an external referral following friends' job arrivals; and finally, line (1.11e) represents the expected value of receiving an internal referral following friends' job arrivals.

An employed worker's value function is characterized by

$$\rho V_1(a,w,p,\Gamma) = U(w) \tag{1.12a}$$

$$+ \delta[V_0(a,\Gamma) - V_1(a,w,p,\Gamma)] \tag{1.12b}$$

$$+ \lambda_1 \int_{q(a,w,p,\Gamma)}^{p} [V_1(a,\phi_1(a,x,p,\Gamma),p,\Gamma) - V_1(a,w,p,\Gamma)]dF(x) \tag{1.12c}$$

$$+ \lambda_1 \int_{p}^{p^{max}} [V_1(a,\phi_1(a,p,x,\Gamma),x,\Gamma) - V_1(a,w,p,\Gamma)]dF(x) \tag{1.12d}$$

$$+ \delta\pi_0 \int_{q(a,w,p,\Gamma)}^{p} [V_1(a,\phi_1(a,y,p,\Gamma),p,\Gamma) - V_1(a,w,p,\Gamma)]d\Gamma_1(y) \tag{1.12e}$$

$$+ \delta\pi_0 \int_{p}^{p^{max}} [V_1(a,\phi_1(a,p,y,\Gamma),y,\Gamma) - V_1(a,w,p,\Gamma)]d\Gamma_1(y) \tag{1.12f}$$

$$+ \lambda_1\pi_1 \int_{q(a,w,p,\Gamma)}^{p} \int_{q(a,w,p,\Gamma)}^{y} [V_1(a,\phi_1(a,x,p,\Gamma),p,\Gamma) - V_1(a,w,p,\Gamma)]dF(x)d\Gamma_1(y) \tag{1.12g}$$

$$+ \lambda_1\pi_1 \int_{p}^{p^{max}} \int_{q(a,w,p,\Gamma)}^{p} [V_1(a,\phi_1(a,x,p,\Gamma),p,\Gamma) - V_1(a,w,p,\Gamma)]dF(x)d\Gamma_1(y) \tag{1.12h}$$

$$+ \lambda_1\pi_1 \int_{q(a,w,p,\Gamma)}^{p} \int_{y}^{p^{max}} [V_1(a,\phi_1(a,y,p,\Gamma),p,\Gamma) - V_1(a,w,p,\Gamma)]dF(x)d\Gamma_1(y) \tag{1.12i}$$

$$+ \lambda_1\pi_1 \int_{p}^{p^{max}} \int_{p}^{y} [V_1(a,\phi_1(a,p,x,\Gamma),x,\Gamma) - V_1(a,w,p,\Gamma)]dF(x)d\Gamma_1(y) \tag{1.12j}$$

$$+ \lambda_1\pi_1 \int_{p}^{p^{max}} \int_{y}^{p^{max}} [V_1(a,\phi_1(a,p,y,\Gamma),y,\Gamma) - V_1(a,w,p,\Gamma)]dF(x)d\Gamma_1(y), \tag{1.12k}$$

where $\phi_1(a,p,p',\Gamma)$ is the bargained wage defined in (1.7), and $q(a,w,p,\Gamma)$ is the productivity threshold defined in (1.9). To understand expression (1.12), note that line (1.12a) represents the flow utility; line (1.12b) represents the expected value of experiencing a job separation; line (1.12c) represents the expected value of a wage raise resulting from a direct job offer; line (1.12d) represents the expected value of a job change resulting from a direct job offer; line (1.12e) represents the expected value of a wage raise resulting from an internal referral following friends' job separations; line (1.12f) represents the expected value of a job change resulting from an internal referral following friends' job separations; lines (1.12g) and (1.12h) represent the expected value of a wage raise resulting from an external referral following friends' job arrivals for friends with productivity lower than and higher than p respectively; line (1.12i) represents the expected value of a wage raise resulting from an internal referral following friends' job arrivals; line (1.12j) represents the expected value of a job change resulting from an external referral following friends' job arrivals; and finally, line (1.12k) represents the expected value of a job change resulting from an internal referral following friends' job arrivals.

In Appendix 3.1, I provide details on solving the reservation productivity $p_0(\cdot)$ and bargained wages $\phi_0(\cdot), \phi_1(\cdot)$. Specifically, I show that for CRRA utility function

$$U(x) = \begin{cases} \ln x & \text{if } \alpha = 1 \\ \frac{x^{1-\alpha}-1}{1-\alpha} & \text{if } \alpha \neq 1 \end{cases}, \tag{1.13}$$

the reservation productivity is given by the implicit function

$$
\ln p_0(a,\Gamma) = \begin{cases}
\ln b \\
+\beta(\lambda_0-\lambda_1)\int_{p_0(a,\Gamma)}^{p^{max}} \frac{\bar{F}(x)x^{-1}}{\rho+\delta+\lambda_1\beta\bar{F}(x)+[\lambda_1\pi_1\beta\bar{F}(x)+\delta\pi_0\beta]\bar{\Gamma}_1(x)}dx \\
+\delta\pi_0\beta\int_{p_0(a,\Gamma)}^{p^{max}} \frac{(\bar{\Gamma}_0(y)-\bar{\Gamma}_1(y))y^{-1}}{\rho+\delta+\lambda_1\beta\bar{F}(y)+[\lambda_1\pi_1\beta\bar{F}(y)+\delta\pi_0\beta]\bar{\Gamma}_1(y)}dy & \text{if } \alpha = 1 \\
+\lambda_1\pi_1\beta\int_{p_0(a,\Gamma)}^{p^{max}} \frac{(\bar{\Gamma}_0(x)-\bar{\Gamma}_1(x))\bar{F}(x)x^{-1}}{\rho+\delta+\lambda_1\beta\bar{F}(x)+[\lambda_1\pi_1\beta\bar{F}(x)+\delta\pi_0\beta]\bar{\Gamma}_1(x)}dx \\
\\
\frac{1}{1-\alpha}\ln\Big\{b^{1-\alpha} \\
+(1-\alpha)\beta(\lambda_0-\lambda_1)\int_{p_0(a,\Gamma)}^{p^{max}} \frac{\bar{F}(x)x^{-\alpha}}{\rho+\delta+\lambda_1\beta\bar{F}(x)+[\lambda_1\pi_1\beta\bar{F}(x)+\delta\pi_0\beta]\bar{\Gamma}_1(x)}dx & \text{if } \alpha \neq 1, \\
+(1-\alpha)\delta\pi_0\beta\int_{p_0(a,\Gamma)}^{p^{max}} \frac{(\bar{\Gamma}_0(y)-\bar{\Gamma}_1(y))y^{-\alpha}}{\rho+\delta+\lambda_1\beta\bar{F}(y)+[\lambda_1\pi_1\beta\bar{F}(y)+\delta\pi_0\beta]\bar{\Gamma}_1(y)}dy \\
+(1-\alpha)\lambda_1\pi_1\beta\int_{p_0(a,\Gamma)}^{p^{max}} \frac{(\bar{\Gamma}_0(x)-\bar{\Gamma}_1(x))\bar{F}(x)x^{-\alpha}}{\rho+\delta+\lambda_1\beta\bar{F}(x)+[\lambda_1\pi_1\beta\bar{F}(x)+\delta\pi_0\beta]\bar{\Gamma}_1(x)}dx\Big\}
\end{cases} \tag{1.14}
$$

and the bargained wages are given by

$$\ln \phi_1(a, p^L, p^H, \Gamma) = \begin{cases} \ln a + \beta \ln p^H + (1-\beta)\ln p^L \\ -\delta\pi_0(1-\beta)^2 \int_{p^L}^{p^H} \frac{\bar{\Gamma}_1(y)y^{-1}}{\rho+\delta+\lambda_1\beta\bar{F}(y)+[\lambda_1\pi_1\beta\bar{F}(y)+\delta\pi_0\beta]\bar{\Gamma}_1(y)}dy \\ -\lambda_1(1-\beta)^2\Big[\int_{p^L}^{p^H} \frac{\bar{F}(x)x^{-1}}{\rho+\delta+\lambda_1\beta\bar{F}(x)+[\lambda_1\pi_1\beta\bar{F}(x)+\delta\pi_0\beta]\bar{\Gamma}_1(x)}dx & \text{if } \alpha = 1 \\ +\pi_1\int_{p^H}^{p^{max}}\int_{p^L}^{p^H} \frac{\bar{F}(x)x^{-1}}{\rho+\delta+\lambda_1\beta\bar{F}(x)+[\lambda_1\pi_1\beta\bar{F}(x)+\delta\pi_0\beta]\bar{\Gamma}_1(x)}dxd\Gamma_1(y) \\ +\pi_1\int_{p^L}^{p^H}\int_{p^L}^{y} \frac{\bar{F}(x)x^{-1}}{\rho+\delta+\lambda_1\beta\bar{F}(x)+[\lambda_1\pi_1\beta\bar{F}(x)+\delta\pi_0\beta]\bar{\Gamma}_1(x)}dxd\Gamma_1(y)\Big] \\ \\ \ln a + \frac{1}{1-\alpha}\ln\Big\{\beta(P^H)^{1-\alpha} + (1-\beta)(P^L)^{1-\alpha} \\ -(1-\alpha)\delta\pi_0(1-\beta)^2 \int_{p^L}^{p^H} \frac{\bar{\Gamma}_1(y)y^{-\alpha}}{\rho+\delta+\lambda_1\beta\bar{F}(y)+[\lambda_1\pi_1\beta\bar{F}(y)+\delta\pi_0\beta]\bar{\Gamma}_1(y)}dy \\ -(1-\alpha)\lambda_1(1-\beta)^2\Big[\int_{p^L}^{p^H} \frac{\bar{F}(x)x^{-\alpha}}{\rho+\delta+\lambda_1\beta\bar{F}(x)+[\lambda_1\pi_1\beta\bar{F}(x)+\delta\pi_0\beta]\bar{\Gamma}_1(x)}dx & \text{if } \alpha \neq 1, \\ +\pi_1\int_{p^H}^{p^{max}}\int_{p^L}^{p^H} \frac{\bar{F}(x)x^{-\alpha}}{\rho+\delta+\lambda_1\beta\bar{F}(x)+[\lambda_1\pi_1\beta\bar{F}(x)+\delta\pi_0\beta]\bar{\Gamma}_1(x)}dxd\Gamma_1(y) \\ +\pi_1\int_{p^L}^{p^H}\int_{p^L}^{y} \frac{\bar{F}(x)x^{-\alpha}}{\rho+\delta+\lambda_1\beta\bar{F}(x)+[\lambda_1\pi_1\beta\bar{F}(x)+\delta\pi_0\beta]\bar{\Gamma}_1(x)}dxd\Gamma_1(y)\Big]\Big\} \end{cases} \tag{1.15}$$

and

$$\phi_0(a, p, \Gamma) = \phi_1(a, p_0(a, \Gamma), p, \Gamma). \tag{1.16}$$

Model Property.

Proposition 1. Job transitions are independent of ability.

All the job transitions, including UE, EE, and EU, are independent of workers' abilities. First, by model assumption, direct job arrivals and job separations are independent of workers' abilities. Additionally, workers' optimal accept/reject decisions are independent of abilities: for employed workers, they always climb the productivity ladder; for unemployed workers, the reservation productivities are independent of their abilities as shown in (1.14). Therefore, referrals are independent of workers' abilities because they are generated as results of employed workers' job separations and accept/reject decisions after direct job offers. Finally, workers' optimal accept/reject decisions with respect to referrals are independent of abilities.

This property is mainly driven by two modeling assumptions. The first assumption is that direct job offer arrivals and job separations do not depend on workers' abilities. This assumption rules out the possibility for high-ability workers to encounter more frequent or better quality jobs, because previous empirical studies find little evidence of sorting between executive ability and firm productivity.[21] This assumption also rules out the possibility for low-ability workers to experience more frequent job separations, because in the executive market most of the separations are voluntary due to personal reasons/retirement/death and only a small fraction are forced due to low competence.[22]

The second assumption is no information asymmetry in wage bargaining. This assumption guarantees that worker ability is fully compensated in all bargained wages, and thus when contacted by the same firm, low-ability and high-ability workers have the same accept/reject incentives. It abstracts from asymmetric information problems because they are not of first order importance for the purpose of studying network effects of referrals. Moreover, in the executive market, the problem of asymmetric information is less severe because there are typically abundant records on a candidate's past performance, upon which a firm can evaluate his/her ability.

[21]For example, Terviö (2008) and Gabaix and Landier (2008) estimate assignment models between firms and CEOs. Under the assumption of positive assortative matching, their empirical results show very small dispersion in CEO ability. This suggests that, quantitatively, there is no significant sorting between worker ability and firm productivity.

[22]For example, Huson et al. (2001), Kaplan and Minton (2006), and Taylor (2010) use the news on the *Wall Street Journal* to categorize whether a separation is voluntary or forced. They show that on average, only 2% of the CEOs are forced to leave their job each year. Additionally, some of the forced separations are caused by personal scandals that are not directly related to the executives' competence in generating output, the ability in this model.

Corollary 1. The speed of career advancement depends on referrals.

Career advancement (UE and EE transitions) depends on referrals because referrals are an additional channel of job arrivals. As discussed earlier, the arrival rate and distribution of referrals are determined by Γ, the productivities of friends' jobs. Variation in Γ leads to variation in referrals and ultimately to variation in the speed of career advancement.

Proposition 2. The level of compensation depends on worker ability.

Specifically, log wage is additively separable in ability as shown in (1.15) and (1.16): $\ln \phi_1(a, p^L, p^H, \Gamma) = \ln a + \ln \phi_1(1, p^L, p^H, \Gamma)$, $\ln \phi_0(a, p, \Gamma) = \ln a + \ln \phi_0(1, p, \Gamma)$. In this model, ability only comes into play in the wage bargaining process, and thus it is reflected in the level of wage.

This property is mainly driven by two parametric assumptions. First, output and unemployment income are multiplicative in worker ability. Second, workers receive their reservation value and a share of the additional rent through bargaining. These are standard assumptions in the literature, with a reasonable depiction of the reality and the convenience of tractability.

It is useful to note that the additive separability of log wage implies that the wage growth rate is independent of ability. Instead, wage growth depends on referrals because it results from competing offers.

Discussion: Networks and Referrals

My model generates a rich set of network effects, and I discuss some of them in this section.

First, a worker's number of friends is not the only thing that matters. A worker's friends act like filters in sending referrals. The distribution of referrals from one particular friend is censored by the productivity of this friend's job. In other words, the best referral one can expect from a friend is as good as the friend's own job. Therefore, the better a friend's job, the better the referral he/she sends. Referral prospects depend not only on the number of friends, but also on friends' success in their careers.

Second, friends of a friend matter. On the one hand, they "compete" for job referrals, which lowers a worker's probability of receiving a particular referral (competition effect).[23] On the other hand, they send referrals to a worker's friend, improving the friend's job. This increases the quantity and the quality of the referrals the friend sends out, which ultimately benefits the worker (ripple effect). These two mechanisms work in opposite directions, and the net effect of friends of friends is qualitatively ambiguous. Through a quantitative analysis with a counterfactual experiment in Section 1.6, I find that the competition effect is more important.

Third, mutual friends matter. More precisely, the clustering structure, i.e., whether a worker's friends are also friends with one another, affects referrals. I illustrate the effects of clustering using Figure 1.1 as an example. Consider worker A in these two networks.[24] According to the definition in (1.3), worker A's local clustering coefficient $c_A = 0$ for the network on the left because none of his/her friends are connected; it is $c_A = \frac{2}{\binom{4}{2}} = \frac{1}{3}$ for the network on the right because there exist two connections among his/her friends over six possible connections. Therefore, worker A's network

[23]Workers do not actively compete for referrals. I use "compete" in a statistical sense.

[24]Note that in both networks, worker A has four friends, and each friend has two friends (a fifty-percent chance for A to receive a friend's referral). Therefore, for worker A, both the number of friends and the level of competition for referral are the same in the two networks. The major difference of the two networks is clustering.

on the right has a higher level of clustering. An advantage of clustering is that it

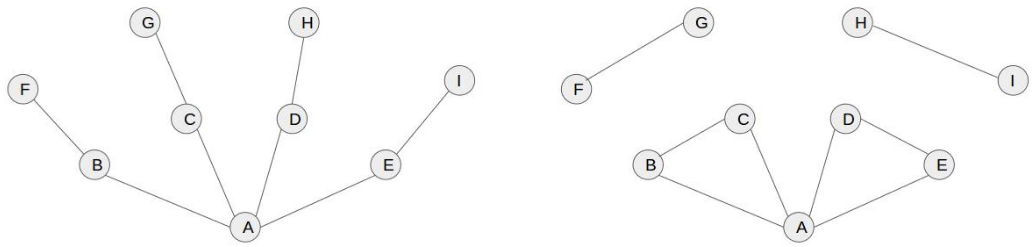

Figure 1.1: Low vs. High Clustering

keeps the positive spillovers in an inner circle (closeness effect). Consider the scenario when friend B sends out a referral, but not to A. In the low-clustering network, this referral goes to F, who will not be able to help A directly. In comparison, in the high-clustering network, this referral goes to C who is a friend of worker A and will be able to help A directly. In this sense, clustering helps. A disadvantage of clustering is that it limits positive spillovers from afar (isolation effect). Consider an alternative scenario when friends $\{B, C, D, E\}$ get unlucky and experience few good shocks. In the low-clustering network, worker A can still benefit in the long run from the good shocks to friends of friends $\{F, G, H, I\}$ through their referrals to $\{B, C, D, E\}$. In the high-clustering network, however, there is no such channel. In this sense, clustering hurts. These two mechanisms work in opposite directions, and the net effect of clustering is qualitatively ambiguous. Through a quantitative analysis with a counterfactual experiment in Section 1.6, I find that the relative importance of these two effects is heterogeneous.

1.3 Data

In this section I describe my data sets and present summary statistics and reduced-form evidence. I combine three data sets: (1) Compustat Executive Compensation (ExecuComp); (2) BoardEx; and (3) the Center for Research in Security Prices (CRSP) U.S. Stock Database.

Data Sets

Compustat Executive Compensation. Compustat Executive Compensation (hereafter, ExecuComp) provides executive compensation data collected directly from companies' annual proxy statements filed with the U.S. Securities and Exchange Commission (SEC).[25] Most companies report around 5 executives for a given year, typically the C-Suite, including the chief executive officer (CEO), the chief financial officer (CFO), and the chief operating officer (COO), etc. I use ExecuComp to construct a panel of individuals' employment history in the executive market along with their on-the-job compensation.

BoardEx. BoardEx provides network connection data among board members and senior executives in notable public and private companies collected from publicly available information. Social connections are defined by overlaps in education, private and public companies, and other social activities such as charities, clubs, business associations, and university board memberships, etc. I use BoardEx to construct three evolving networks for education, workplace, and social activities respectively. Addi-

[25] A proxy statement is a document containing the information the Securities and Exchange Commission (SEC) requires companies to provide to shareholders so they can make informed decisions about matters that will be brought up at an annual or special stockholder meeting. It is also known as Form DEF 14A.

tionally, BoardEx also provides information on the gender and nationality composition of the board of the directors.

U.S. Stock Database from the Center for Research in Security Prices. The U.S. Stock Database from the Center for Research in Security Prices (hereafter, CRSP) provides stock market data for equity securities traded on the major U.S. stock exchanges. CRSP contains information such as price, quote, market capitalization, shares outstanding, trading volume, etc. I use CRSP to construct companies' financial variables.

Sample Construction. In my estimation sample, I include individuals who have ever worked as senior executives in S&P 500 companies between (and including) 2007 and 2015 and merge the data in ExecuComp and BoardEx.[26] The resulting sample consists of a total of 4,192 individuals. In constructing employment history, I code the time when an individual is not an executive in an S&P 500 firm as "unemployment". This should not be interpreted literally; instead it should be interpreted as not being employed in the specific labor market under study. Additionally, as my model rules out downward job-to-job transition, I code a transition from an executive job in a high market cap company to another executive job in a low market cap company as if the individual goes through unemployment in between the jobs. Finally, I construct three evolving social networks. For each year t, a network includes all the connections established before t. Additionally, I assume that once a connection is formed, it will

[26] I use a sample period after 2006 because of a change in the reporting rule of the executives' total compensation. In 2006, the Financial Accounting Standard (FAS) 123R changed the reporting requirements of the DEF 14A form. Under this new reporting regime, the cost of all employee stock options, as well as other equity-based compensation arrangements, have to be reflected in the financial statements based on the estimated fair value of the awards.

never be lost. The resulting education network is fixed over time, and the work and social-activity networks monotonically grow larger over time. It is useful to note that not all work connections are established as a result of movements in the executive labor market. Examples of exogenous work connections are those formed when one or both individuals serve on the board of directors or in lower management positions. In the end, I discuss the potential issue of missing connections in the constructed networks. On the one hand, connections with individuals out of the sample are not included. On the other hand, some connections between individuals in the sample may not be recorded in BoardEx if the information is not publicly available such as attending the same church. Due to these data limitations, the constructed networks are subnetworks of the "true" networks and may potentially lead to underestimation of the effect of referrals.

Summary Statistics

Executives and Firms. Table 1.1 reports the summary statistics for the executives and the firms in the data. Executives in S&P 500 firms have an average annual compensation of 6.09 million dollars, with a median of 3.89 million. Most of the individuals in the sample are between the ages of 38 and 61 at the beginning of the sample period, with both the mean and the median being around 49. Moreover, 91% of them are male. S&P 500 firms in the sample period have an average monthly stock return of 0.69% and an average market capitalization of 29.47 billion dollars. The size of the board of directors ranges from 7 to 15, with a mean of 10.96. The average proportion of male directors is 85%, and the average proportion of foreign directors is 11%.

Variable	Mean	5%	25%	50%	75%	95%
			Panel A: Executives			
Compensation	6.09	1.15	2.31	3.89	7.27	17.68
Age in 2007	48.98	38	44	49	54	61
Male	0.91					
			Panel B: Firms			
Stock Return	0.69%	-0.62%	0.35%	0.71%	1.11%	2.21%
Market Capitalization	29,472.24	4,698.01	7,481.18	13,489.17	29,597.99	121,726.30
Board Size	10.96	7.80	9.31	10.87	12.18	14.81
Board Gender Ratio	0.85	0.72	0.81	0.85	0.89	0.97
Board Nationality Mix	0.11	0	0	0.07	0.18	0.38

Table 1.1: Summary Statistics: Executives and Firms

Notes: (1). Compensation is the the total annual compensation reported in the SEC filings. It is the sum of the salary, bonus, the value of stock and option awards, non-equity incentive plan compensation, change in pension value and nonqualified deferred compensation earnings, and other compensations such as perquisites and personal benefits etc. Compensation reported in this table is adjusted for inflation, in the unit of 2015 U.S. million dollar. (2). Stock return is monthly return. (3). Market capitalization is adjusted for inflation, in the unit of 2015 U.S. million dollar. (4). Board size is the number of directors on the board. (5). Board gender ratio is the proportion of male directors. (6). Board nationality mix is the proportion of foreign directors.

Networks. Table 1.2 reports the summary statistics for the social networks. For each network, I report the distributions of degree (defined in (1.2)) and clustering coefficient (defined in (1.3)) at the beginning (2007), the middle (2011), and the end (2015) of the sample period. The education network does not change over time. About half of the individuals have no school friends. Conditional on having school friends, the average number of school friends is 6.35. The work network grows moderately; the average number of work friends increases from 16.40 in 2007 to 24.76 in 2015. The social-activity network grows more rapidly; the average number of social-activity friends increases from 0.45 in 2007 to 13.57 in 2015. These new social-activity connections are formed unevenly. About half of the individuals have no social-activity connections throughout the entire sample period, whereas the top 5% added more than 60 new connections.

These networks exhibit two key features. The first feature is high level of clustering.[27] The average clustering coefficient conditional on $d > 0$ ranges from 0.40 to 0.65, meaning for an average individual more than 40% of his/her friends are also friends with one another.[28] In comparison, the average clustering coefficient for a randomly formed network is around 0.05. The second feature is sorting on degree, i.e., high-degree individuals' friends are often high-degree, and low-degree individuals' friends are often low-degree. The correlations between own degree and the average of friends' degrees are 0.85, 0.69, and 0.74 respectively for the education, work, and social-activity networks. In comparison, the correlation for a randomly formed network is around 0.20.

Labor Market Transitions. Tables 1.3 and 1.4 report the summary statistics for the labor market transitions. Each year, the average fraction of non-executives is 47.21%, and their annual transition rate from non-executive to executives (UE) is 16.09%. The average fraction of executives is 52.79%, their annual transition rate to executives in a more productive firm (EE) is 0.72%, and their annual transition rate to non-executives (EU) is 13.03%. Career advancement is defined as either a UE or EE transition. The annual rate of career advancement is 7.97%. The average duration of the non-executive spell is 3.97 years for the right-censored spells and 3.87 years for those ending with an executive job. The average duration of an executive job spell is

[27] The definition of a connection requires two individuals to have overlapping time in an organization. Therefore, with people joining and leaving at staggered time, the resulting networks are not collections of cliques (a clique is a subgraph such that every two distinct nodes are connected; all the nodes in a clique have clustering coefficients of 1).

[28] Clustering coefficient is meaningful only for individuals with friends. Clustering coefficient for an individual with no friend is defined to be 0.

Variable	Mean	Mean Conditional on $d > 0$	5%	25%	50%	75%	95%
		Panel A: Network Y^{Edu}					
d^{Edu}	3.67	6.35	0	0	1	5	24
c^{Edu}	0.27	0.47	0	0	0	0.52	1
		Panel B: Network Y^{Work}					
$d^{Work,2007}$	16.40	17.08	1	7	11	20	52
$c^{Work,2007}$	0.63	0.65	0	0.41	0.64	0.92	1
$d^{Work,2011}$	20.76	20.97	5	9	14	25	61
$c^{Work,2011}$	0.60	0.61	0.20	0.39	0.58	0.86	1
$d^{Work,2015}$	24.76	24.76	7	11	18	31	69
$c^{Work,2015}$	0.56	0.56	0.21	0.35	0.52	0.77	1
		Panel C: Network Y^{Social}					
$d^{Social,2007}$	0.45	3.87	0	0	0	0	3
$c^{Social,2007}$	0.04	0.40	0	0	0	0	0.41
$d^{Social,2011}$	4.39	12.25	0	0	0	2	26
$c^{Social,2011}$	0.18	0.52	0	0	0	0.32	0.89
$d^{Social,2015}$	13.57	26.48	0	0	1	9	66
$c^{Social,2015}$	0.27	0.53	0	0	0	0.53	0.96
		Panel D: Network Y^{All}					
$d^{All,2007}$	20.42	20.92	2	8	15	26	60
$c^{All,2007}$	0.50	0.51	0.13	0.29	0.43	0.69	1
$d^{All,2011}$	28.41	28.57	5	12	20	37	78
$c^{All,2011}$	0.46	0.46	0.16	0.27	0.39	0.61	1
$d^{All,2015}$	41.15	41.15	8	15	27	49	130
$c^{All,2015}$	0.43	0.43	0.17	0.26	0.37	0.55	0.94

Table 1.2: Network Statistics

Notes: (1). d is degree, defined in Eq.(1.2). It is the number of an individual's friends. (2). c is clustering coefficient, defined in Eq.(1.3). It gives the proportion of an individual's friends who are friends with one another. (3). The time superscript for Y^{Edu} is omitted because it does not change in the sample period.

5.26 years for the right-censored spells, 3.30 years for those ending with transitions to non-executives, and 3.10 years for those ending with transitions to executive jobs in more productive firms.

	Conditional on U	Conditional on E		Unconditional
	UE Transition	EE Transition	EU Transition	Career Advancement
Annual Transition Rate	16.09%	0.72%	13.03%	7.97%

Table 1.3: Job Transition

Notes: (1). A UE transition is defined as the transition from non-executive to executive. (2). An EE transition is defined as the transition from one executive job to another. (3). An EU transition is defined as the transition from executive to non-executive. (4). A career advancement is defined as either a UE or EE transition.

Panel A: Unemployment Spell			
Type of Transition	Number of Spells	Mean of Spell Duration	Std. Dev. of Spell Duration
Right Censored	1,984	3.97	2.32
To Employment	2,547	3.87	2.37
Panel B: Job Spell			
Type of Transition	Number of Spells	Mean of Spell Duration	Std. Dev. of Spell Duration
Right Censored	2,208	5.26	3.08
To Unemployment	2,422	3.30	2.15
To Another Job	127	3.10	1.85

Table 1.4: Durations

Notes: (1). An unemployment spell is defined as the period when an individual is not an executive. (2). The unit of time is one year.

Reduced-form Evidence

In this part, I present suggestive evidence on the presence of job referrals in the executive labor market and supporting evidence on my model assumption of no immediate "chain effect" from friends of friends.

Correlation in Friends' Compensation. This analysis shows that socially connected executives' compensation is correlated, which is consistent with the implication of my job referral model. I regress an executive's compensation on the average of his/her friends' compensation and a set of covariates according to the following Spatial Auto-Regressive model (SAR):

$$\ln w_{i,t} = \rho \frac{\sum_{j \in N^{k,t}(i)} \ln w_{j,t}}{d_i^{k,t}} + (x_{i,t})^T \beta + u_i + v_t + \epsilon_{i,t}, \tag{1.17}$$

where $\ln w_{i,t}$ is executive i's log compensation in year t, and $\frac{\sum_{j \in N^{k,t}(i)} \ln w_{j,t}}{d_i^{k,t}}$ is the average of i's executive friends' log compensation in year t. $x_{i,t}$ is a vector of exogenous covariates including the executive's age, his/her employer's financial variables and governance variables, and industry dummy. u_i is the individual fixed-effects, and v_t is the time fixed-effects. $\epsilon_{i,t}$ is an i.i.d. error term. Coefficient ρ, sometimes referred to as the "social interaction parameter", measures the percentage change in i's compensation in response to a one percentage increase to friends' compensation.

In this regression, $\frac{\sum_{j \in N^{k,t}(i)} \ln w_{j,t}}{d_i^{k,t}}$ is correlated with the error term $\epsilon_{i,t}$ because of simultaneous equations. To address the endogeneity problem, I use instrumental variables. The instruments are the exogenous covariates of friends and friends of friends. More precisely, (1.17) can be written in the following compact form

$$\ln w = \rho W \ln w + x\beta + u + v + \epsilon, \tag{1.18}$$

and I use Wx and $W^2 x$ as instruments for $W \ln w$.

Table 1.5 reports the results of the Spatial Auto-Regressive regression in (1.17), where different columns correspond to different network specifications. Column (1) uses the simplified network, which does not distinguish between different types of social connections; column (2) uses the school network; column (3) uses the work network;

and column (4) uses the social-activity network. These results illustrate that, across all specifications of networks, an executive's compensation is positively correlated with his/her friends' compensation. This correlation is driven by neither the correlation in time-invariant individual characteristics nor time- or industry-specific shocks because of the inclusion of individual, time, and industry fixed-effects. The positive correlation in socially connected executives' compensation supports my model's prediction that executives with better jobs send better quality referrals, increasing their friends' compensation.

	Dependent variable: log(Comp)			
	(1)	(2)	(3)	(4)
Mean (Friend's log(Comp))	0.3858***			
	(0.0274)			
Mean (School Friend's log(Comp))		0.2105***		
		(0.0404)		
Mean (Work Friend's log(Comp))			0.5573***	
			(0.0364)	
Mean (Social-Activity Friend's log(Comp))				0.0207*
				(0.0082)
Observations	18,481	18,481	18,481	18,481
Year FE	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes

Table 1.5: Correlation of Socially Connected Executives' Compensation

Notes: (1). This table summarizes the results of a Spatial Auto-Regressive regression with individual, year, and industry fixed effects, where the dependent variable is the log of annual compensation. (2). Column (1) uses the simplified network, which does not distinguish between types of social connections; column (2) uses the school network; column (3) uses the work network; and column (4) uses the social-activity network. (3). Instrumental variables are used to address the endogeneity problem arising from simultaneous equations. (4). Other covariates include age, age^2, and dummy for serving on the board of directors; the board's size, gender ratio, and nationality mix; and firm's lagged log(market cap) and stock return. (5). Standard errors are reported in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Co-movement in Friends' Job Transitions. This analysis shows that individuals are more likely to make career advancement when their executive friends leave their current jobs, which suggests the presence of referrals. Table 1.6 reports the results of a Logit regression. It shows that an individual is more likely to make an upward transition (UE or EE) when there is a friend(s) who experiences job-to-job transition (EE) or job separation (EU) in the same or the previous year. This corresponds to internal referrals following direct job arrivals or job separations. The individual fixed effects control for time-invariant individual characteristics, and the year fixed effects rule out time-specific shocks.[29] It is useful to note that given the small number of EE transitions, most of the co-movements are in the form of individuals making upward transitions following friends' downward transitions, which are unlikely to be driven by cohort effect or industry-specific shocks. To interpret the quantitative results, note that the coefficients in Logit regressions represent log odds ratios. For example, a log odds ratio of 0.6957 is equivalent to an odds ratio of 2.005, meaning the odds of career advancement for an individual whose friends change jobs is twice that of an individual whose friends do not change jobs. In terms of the probability of career advancement, it corresponds to a 4.7% increase in absolute terms according to the OLS estimates in Table 3.2.

Lack of Co-movement Beyond Immediate Friends. This analysis shows the lack of co-movement in job transitions beyond immediate friends, which motivates my model assumption ruling out immediate "chain effect" from friends of friends. Table

[29]The Logit regression with individual fixed effects uses a subsample of the individuals with the same total number of career advancements: in this case, individuals with exactly one career advancement during the sample period. As a robustness check, the results of an OLS regression with year and individual fixed effects using the full sample are reported in Table 3.2.

Dependent variable: $\mathbb{1}$*(Career Advancement)*$\in \{0,1\}$				
	(1)	(2)	(3)	(4)
$\mathbb{1}$ (Job Transition				
Among Executive Friends)	0.6957***			
	(0.0934)			
...from School		0.2706***		
		(0.0769)		
...from Work			0.7954***	
			(0.0870)	
...from Social Activity				0.2508**
				(0.0944)
Observations	19,360	19,360	19,360	19,360
Individual FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes

Table 1.6: Co-movement in Socially Connected Executives' Job Transitions

Notes: (1). This table summarizes the results of a Logit regression with year and individual fixed effects, where the dependent variable is whether an individual experiences a career advancement. (2). A career advancement is defined as either a transition from non-executive job to executive job (UE) or a transition between executive jobs with productivity increase (EE). (3). Variable $\mathbb{1}$ (Job Transition Among Executive Friends) is a dummy variable that equals one if any friend experiences a transition between executive jobs with productivity increase (EE) or a transition from executive job to non-executive job (EU) in the same year or the previous year. (4). The scope of friends varies for different network specifications. Column (1) uses the simplified network, which does not distinguish between types of social connections; column (2) uses the school network; column (3) uses the work network; and column (4) uses the social-activity network. (5). The Logit regression with individual fixed effects uses a subsample of the individuals with the same total number of career advancements: in this case, individuals with exactly one career advancement during the sample period. (6). Other covariates include age^2, age^3, whether an individual was an executive in the previous year, and the productivity of the job in the previous year if the individual was an executive. (7). Standard errors are reported in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

1.7 shows that an individual's probability of career advancement responds to the job-to-job transition (EE) and job separation (EU) of immediate friends, but not to that of friends of friends.[30] The results suggest that, in the short run, individuals do not significantly benefit from friends of friends.

[30] I exclude immediate friends in defining friends of friends and only include those who are not immediate friends. Additionally, in this exercise, I use the simplified network Y^{All} which does not distinguish between different types of social connections because it is problematic to restrict to friends of friends in each of the networks $Y^{Edu}, Y^{Work}, Y^{Social}$. Such restriction excludes a significant number of friends of friends across different types of social connections, for example, a work friend of a school friend.

Dependent variable: $\mathbb{1}$*(Career Advancement)*$\in \{0,1\}$		
	(1)	(2)
$\mathbb{1}$(Job Transition Among Executive Friends)	0.6338***	
	(0.1401)	
$\mathbb{1}$(Job Transition Among Executive Friends of Friends)	0.2637	
	(0.4536)	
Number of Job Transition Among Executive Friends		0.0595**
		(0.0216)
Number of Job Transition Among Executive Friends of Friends		0.0035
		(0.0031)
Observations	19,360	19,360
Individual FE	Yes	Yes
Year FE	Yes	Yes

Table 1.7: Lack of Co-movement in Job Transition Beyond Immediate Friends

Notes: (1). This table summarizes the results of a Logit regression with year and individual fixed effects, where the dependent variable is whether an individual experiences a career advancement. (2). A career advancement is defined as either a transition from non-executive job to executive job (UE) or a transition between executive jobs with productivity increase (EE). (3). The simplified network, which does not distinguish between types of social connections, is used to define friends as well as friends of friends. (4). Friends of friends exclude immediate friends and only include those who are not immediate friends. (5). Variable $\mathbb{1}$(Job Transition Among Executive Friends (or Friends of Friends)) is a dummy variable that equals one if any friend (or friend of friends) experiences a transition between executive jobs with productivity increase (EE) or a transition from executive job to non-executive job (EU) in the same year or the previous year. (6). Variable Number of Job Transition Among Friends (of Friends of Friends) is the total number of EE or EU transitions among friends (or friends of friends) in the same year or the previous year. (7). The Logit regression with individual fixed effects uses a subsample of the individuals with the same total number of career advancements: in this case, individuals with exactly one career advancement during the sample period. (8). Other covariates include age^2, age^3, whether an individual was an executive in the previous year, and the productivity of the job in the previous year if the individual was an executive. (9). Standard errors are reported in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

1.4 Estimation Strategy

In this section I present my estimation method. In my empirical application, I use the following specifications. The productivity distribution upon job arrival, F, is truncated log normal:

$$\ln p \sim \mathcal{N}(\mu_p, \sigma_p), \quad p \in [p^{min}, p^{max}]. \tag{1.19}$$

Following Terviö (2008) and Gabaix and Landier (2008), I use the firm's market capitalization as the empirical counterpart for productivity. Additionally, I set p^{min} and p^{max} to be the observed lowest and highest market capitalization. An executive's bargaining power is a function of his/her number of executive friends:

$$\beta = \frac{\exp(\beta_0 + \beta_1 d^{All\cap e})}{1 + \exp(\beta_0 + \beta_1 d^{All\cap e})}, \tag{1.20}$$

where $d_i^{All\cap e} = |\{j | j \in N^{All}(i), s_j = e\}|$ is i's number of executive friends. Executive ability a follows a log normal distribution:

$$\ln a \sim \mathcal{N}(\mu_a, \sigma_a). \tag{1.21}$$

Log compensation has additive i.i.d. measurement error ϵ:

$$\epsilon \sim \mathcal{N}(0, \sigma_\epsilon).^{31} \tag{1.22}$$

The unit of time is one year, and I set the continuous time discount rate ρ to be 0.05 (an annual discount rate of 0.95).[32]

---

[31] This measurement error should be interpreted as benefits in the compensation package that are not included in the calculation of the value of total compensation. For example, perks such as traveling with a private jet.

[32] Flinn and Heckman (1982) shows that it is difficult to separately identify the discount factor from the flow unemployed income in standard search models.

I estimate the parameters via a minimum-distance estimator, which follows Imbens and Lancaster (1994) and Petrin (2002). The objective function is based on the generalized method of moments (GMM) with three sets of moments. The first two sets of moments are in the form of log-likelihood, which I aim to maximize by requiring that the first derivatives should equal zero. Specifically, the first set of moments characterizes the likelihood of the executives' compensation immediately after career advancement; the second set of moments characterizes the likelihood of the productivities of socially isolated workers' first jobs. The third set of moments are in the form of traditional moments, characterizing executives' labor market transitions. Accordingly, parameters $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)$ can be grouped into three sets, where $\theta_1 = \langle \alpha, \beta_0, \beta_1, \ln(b), \sigma_a, \sigma_\epsilon \rangle$ is mostly relevant to the first set of moments, $\theta_2 = \langle \mu_p, \sigma_p \rangle$ is most relevant to the second set, and $\theta_3 = \langle \lambda_0, \lambda_1, \delta, \pi_0, \pi_1, \nu_u, \omega^{Edu}, \omega^{Work} \rangle$ is most relevant to the third set.

Specifically, the targeted moments are

$$\mathcal{M}(\boldsymbol{\theta}) = \begin{bmatrix} \frac{\partial \ln(L_1(\boldsymbol{\theta}))}{\partial \boldsymbol{\theta_1}} \\ \frac{\partial \ln(L_2(\boldsymbol{\theta}))}{\partial \boldsymbol{\theta_2}} \\ \boldsymbol{m} - \mathbf{E}[\boldsymbol{m}; \boldsymbol{\theta}] \end{bmatrix}, \tag{1.23}$$

where L_1 is the first likelihood of executive compensation, which I discuss in more detail in Section 1.4; and L_2 is the second likelihood of firm productivity, which I discuss in more detail in Section 1.4; and $\boldsymbol{m}$ is a vector of moments for labor market transitions, which I describe in Section 1.4. Then I estimate the parameters using the following objective function:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \mathcal{M}(\boldsymbol{\theta})' W \mathcal{M}(\boldsymbol{\theta}), \tag{1.24}$$

where W is the weight matrix. I choose W to be the diagonal elements of the optimal weight matrix.[33]

Likelihood of Executive Compensation

L_1 is the likelihood of the executives' compensation immediately after career advancement. When executive i experiences a career advancement (UE or EE transition) at time t, the observed compensation in the new job is given by

$$\ln w_{it} = \ln \phi_1(1, p_{it}^L, p_{it}^H, \Gamma_{it}) + \ln a_i + \epsilon_{it}, \tag{1.25}$$

where p_{it}^H is the productivity of the new job, p_{it}^L is the productivity of the old job for an EE transition or the reservation productivity calculated according to (1.14) for a UE transition, ϕ_1 is the equilibrium wage function (1.15), a_i is unobserved ability, and ϵ_{it} is unobserved measurement error. Let $u_i = \ln a_i - \mu_a$, then log compensation can be rewritten as

$$\ln w_{it} = \ln \phi_1(1, p_{it}^L, p_{it}^H, \Gamma_{it}) + \mu_a + u_i + \epsilon_{it}. \tag{1.26}$$

This is a standard random effect model (Maddala, 1971) satisfying

$$E(u_i | p_{it}^L, p_{it}^H, \Gamma_{it}) = 0, \tag{1.27}$$

$$E(\epsilon_{it} | p_{it}^L, p_{it}^H, \Gamma_{it}) = 0,\ E(u_i \epsilon_{it} | p_{it}^L, p_{it}^H, \Gamma_{it}) = 0. \tag{1.28}$$

The orthogonality condition (1.27) holds because of the model implication that worker i's ability is uncorrelated with his/her own and his/her friends' job dynamics, which is shown in Proposition 1. Condition (1.28) holds because of the independence assumption on the measurement errors.

[33] Altonji and Segal (1996) show that the asymptotic optimal weight matrix may induce bias in finite sample estimates.

Additionally,

$$\mathrm{E}(\mathrm{u}_i u_j | p_{it}^L, p_{it}^H, \Gamma_{it}) = \begin{cases} \sigma_a^2 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}, \quad E(\epsilon_{it}\epsilon_{js} | p_{it}^L, p_{it}^H, \Gamma_{it}) = \begin{cases} \sigma_\epsilon^2 & \text{if } i = j \text{ and } t = s \\ 0 & \text{otherwise} \end{cases}. \tag{1.29}$$

The complete expression of the likelihood function is given in Appendix 3.3.

Likelihood of Firm Productivity

L_2 is the likelihood of the productivities of socially isolated workers' first jobs. The isolated unemployed workers only face direct job arrivals with productivity distribution F. Upon arrival, they accept jobs above their reservation productivity p_0 as described by (1.14). Therefore, their accepted jobs are distributed independently according to F truncated below by p_0. The complete expression of the likelihood function is given in Appendix 3.3.

Moments of Labor Market Transitions

Referrals introduce interdependence in friends' jobs and thus interdependence in labor market transitions, which is difficult to characterize by likelihood. Instead, I simulate the model and target some moments of the transitions.

Targeted Moments. I include the following standard moments: the frequencies of UE, EE, and EU transitions, and the means and the standard deviations of the employment and the unemployment spells. In addition, I include moments characterizing the co-movement in socially connected individuals' job transitions to capture the effects of referrals. These moments include, for each type of social connection, the fraction of unemployed individuals experiencing UE transition conditional on whether they have friends experiencing EU transition, the fraction of unemployed individuals experiencing UE transition conditional on whether they have friends experiencing EE transition, the fraction of employed individuals experiencing EE transition conditional on whether they have friends experiencing EU transition, and the fraction of employed individuals experiencing EE transition conditional on whether they have friends experiencing EE transition.

Simulation. In the model, direct job arrivals and job separations are independent continuous-time Poisson processes.[34] In the simulation, I preserve the continuous-time setting by simulating the waiting times.[35] After the simulation, I discretize the simulated data so that it is comparable to the discrete-time observations in the data. The simulation procedure is described in the following.

The simulation follows an iterative procedure. Each iteration k starts at time t^{k-1} with the employment/unemployment status and job productivity of all the workers being $\mathbf{s} = (s_1, ..., s_n)$ and $\mathbf{p} = (p_1, ...p_n)$. For each unemployed worker, simulate one waiting time for direct job offer $\Delta t_i^{\text{ArrivalU},k}$; for each employed worker, simulate one waiting time for direct job offer $\Delta t_i^{\text{ArrivalE},k}$ and another waiting time for job separation $\Delta t_i^{\text{Separation},k}$. The shortest waiting time among all is $\Delta t^{*,k} = \min\Big\{ \{\Delta t_i^{\text{ArrivalU},k} : s_i = u\}, \{\Delta t_i^{\text{ArrivalE},k}, \Delta t_i^{\text{Separation},k} : s_i = e\} \Big\}$.[36] Set $t^k = t^{k-1} + \Delta t^{*,k}$. If $t^k > T$, where T is

[34] The continuous-time setting rules out simultaneous events.

[35] For a Poisson process with arrival rate λ, at any given time point, the waiting time until the next arrival follows an exponential distribution with mean $\frac{1}{\lambda}$, i.e., the cdf of waiting time is $F(t) = 1 - e^{-\lambda t}$.

[36] In the implementation, I use a slightly different but statistically equivalent approach to reduce computational intensity. Instead of sampling separate arrival times for each person, I only sample three order statistics $t^{\text{ArrivalU}} = \min\{t_i^{\text{ArrivalU}} : s_i = u\}$, $t^{\text{ArrivalE}} = \min\{t_i^{\text{ArrivalE}} : s_i = e\}$, $t^{\text{Separation}} = \min\{t_i^{\text{Separation}} : s_i = e\}$. t^{ArrivalU} follows an exponential distribution with mean $\frac{1}{|\{i:s_i=u\}|\cdot\lambda_0}$;

a pre-specified length of time, the simulation is completed. Otherwise, proceed with the following. Record t^k as the time of the k-th event, along with the corresponding event type $\in$ {Arrival for U, Arrival for E, Separation} and person i^*. If the event is an arrival for an unemployed worker i^*, first sample the job's productivity, then the worker makes the decision to accept or reject (s_{i^*} and p_{i^*} change accordingly). If the event is an arrival for an employed worker i^*, first sample the job's productivity, then the worker makes the decision to accept or reject (s_{i^*} and p_{i^*} change accordingly), which leads to an internal or external referral with probability π_1. If a referral is generated, sample a recipient $j^* \in N^{All,t^k}(i^*)$, then the recipient makes the decision to accept or reject (s_{j^*} and p_{j^*} change accordingly). If the event is a job separation, i^* becomes unemployed (s_{i^*} and p_{i^*} change accordingly), which leads to an internal referral with probability π_0. If a referral is generated, sample a recipient $j^* \in N^{All,t^k}(i^*)$, then the recipient makes the decision to accept or reject (s_{j^*} and p_{j^*} change accordingly). This completes iteration k. The next iteration starts at time t^k with the new $\mathbf{s}$ and $\mathbf{p}$. Note that the memoryless property allows me to "reset the Poisson clock" in each iteration.[37] Specifically, in each iteration, I only use the shortest waiting time and do not carry the rest of the waiting times to the subsequent iterations. For the subsequent iteration, I simulate a new set of waiting times to generate the arrival time and the type of the event.

similarly, t^{ArrivalE} follows an exponential distribution with mean $\frac{1}{|\{i:s_i=e\}|\cdot\lambda_1}$, and $t^{\text{Separation}}$ follows an exponential distribution with mean $\frac{1}{|\{i:s_i=e\}|\cdot\delta}$.

[37]Memoryless refers to the property that the distribution of the waiting time (until an arrival) does not depend on how much time has elapsed already.

Identification

In this part, I provide heuristic arguments for identification of the structural parameters. The first set of parameters, $\theta_1 = \langle \alpha, \beta_0, \beta_1, \ln(b), \sigma_a, \sigma_\epsilon \rangle$, are identified from executive compensation immediately after career advancement. The utility parameter α governs the inter-temporal incentives, and it is therefore identified from the curvature of the log compensation function with respect to log productivity.[38] The bargaining parameters β_0 and β_1 are identified from the overall correlation between log compensation and log productivity, and the intuition is the following. In one extreme case of no bargaining power, an executive's initial compensation is negatively correlated with employer productivity because he/she trades lower present compensation for higher continuation value. In the other extreme case of full bargaining power, an executive's compensation is equal to the marginal product, and thus is positively correlated with employer productivity. Unemployed productivity $\ln(b)$ is identified from the within-person difference between compensation after UE transition and EE transition.[39] The dispersion parameter of ability σ_a is identified from the covariance of within-person residual terms, and the dispersion parameter of measurement error σ_ϵ is identified from the variance of the residual terms net of σ_a^2.

The second set of parameters, the location and dispersion parameters of productivity distribution, $\theta_2 = \langle \mu_p, \sigma_p \rangle$, are identified from the observed distribution of the productivities of socially isolated workers' first jobs. With the log-normal parametric assumption, they are identified from the mean and variance of the observed distribution.

[38] $\frac{1}{\alpha}$ gives the inter-temporal elasticity of substitution.

[39] When $\alpha \neq 1$, $\ln(b)$ and $\mathbb{E}\ln(a)$ can be separately identified even without within-person comparison. The reason is that $\mathbb{E}\ln(a)$ enters the expression of $\ln w$ linearly, while $\ln(b)$ enters non-linearly.

The third set of parameters, $\theta_3 = \langle \lambda_0, \lambda_1, \delta, \pi_0, \pi_1, \nu_u, \omega^{Edu}, \omega^{Work} \rangle$, are identified from the labor market transitions. The job separation rate δ is identified from the duration of employment spell and the frequency of employment-to-unemployment transition. The probability of referral following job separation π_0 is identified from the co-movement of individuals' career advancement and their friends' job separations. Similarly, the probability of referral following job arrival π_1 is identified from the co-movement of individuals' career advancement and their friends' job-to-job transitions. The probability of referring an unemployed friend ν_u is identified by comparing the co-movement of individuals' unemployment-to-employment transitions and their friends' job-to-job transitions or job separations vs. the co-movement of individuals' job-to-job transitions and their friends' job-to-job transitions or job separations, given the productivity distribution. The job arrival rates λ_0 and λ_1 are identified from the frequency of unemployment-to-employment transition and job-to-job transition given the job separation rate δ, the referral parameters (π_0, π_1, ν_u), and the productivity distribution parameters (μ_p, σ_p). The probabilities of referring through different types of social connections, ω^{Edu} and ω^{Work}, are identified from the difference in co-movement pattern in the three different social networks.[40]

1.5 Estimation Results

Parameter Estimates

Table 1.8 reports the parameter estimate. The estimates of the bargaining parameters show that the executives generally have high bargaining power, and their

[40] $\omega^{Social} = 1 - \omega^{Edu} - \omega^{Work}$.

bargaining power is increasing in their number of executive friends. Most of the individuals' bargaining powers range from 0.62 to 0.79.[41] The estimate of the unemployment productivity implies an annual income of 0.52 million dollars for an average-ability individual in a non-executive job. The estimated offer arrival rates for non-executives and executives are 0.130 and 0.006 respectively, implying average waiting times of 7.67 years and 165 years respectively for direct job offers. The estimated referral probabilities after job separation and job arrival are 0.228 and 0.794 respectively. The referral probability after job separation is lower than that after job arrival potentially because some separations are forced, and in these cases there is no chance for referral. The estimated probability of referring non-executive friends is 0.626, and that of referring executive friends is 0.374. Therefore, the model implication that the executives are more selective in accepting jobs, combined with the empirical estimates that they receive fewer direct offers as well as referrals, jointly contribute to the low number of job-to-job transitions observed in the data. Finally, the estimated probabilities of referral through different types of networks show that the executives are most likely to refer their work friends (with probability 0.653), then their social-activity friends (with probability 0.192), and lastly their school friends (with probability 0.155).

Statistical and Economical Significance of Referrals

It is useful to note that the estimated referral probabilities π_0 and π_1 are both statistically and economically significantly different from zero. A Wald test of the joint hypotheses $H_0 : \pi_0 = \pi_1 = 0$ rejects a nested model with no referral, $\chi^2(2) = 1171618$, p=0.[42] To further understand the economic significance of referrals, I decompose the

[41] 99% of the individuals have fewer than 50 *executive* friends.
[42] The critical value for a significance level of 0.001 is 13.82.

Parameter	Notation	Estimate	Std. Err.
Panel A: Parameters in $\theta_1 = \langle \alpha, \beta_0, \beta_1, \ln(b), \sigma_a, \sigma_\epsilon \rangle$			
CRRA Coefficient	α	1.393	0.13340
Constant in the Logit Function for Bargaining Power	β_0	0.500	0.38150
Slope in the Logit Function for Bargaining Power	β_1	0.017	0.00240
Log (Unemployment Productivity)	$\ln(b)$	6.866	0.07653
Dispersion Para. of Executives' Lognormal Ability Dist.	σ_a	0.355	0.00725
Dispersion Para. of Compensations' Normal Measurement Error Dist.	σ_ϵ	0.634	0.00088
Panel B: Parameters in $\theta_2 = \langle \mu_p, \sigma_p \rangle$			
Location Para. of Firms' Lognormal Productivity Dist.	μ_p	9.550	0.01772
Dispersion Para. of Firms' Lognormal Productivity Dist.	σ_p	1.240	0.00083
Panel C: Parameters in $\theta_3 = \langle \lambda_0, \lambda_1, \delta, \pi_0, \pi_1, \nu_u, \omega^{Edu}, \omega^{Work} \rangle$			
Offer Arrival Rate for the Unemployed	λ_0	0.130	0.00004
Offer Arrival Rate for the Employed	λ_1	0.006	0.00001
Job Separation Rate	δ	0.147	0.00004
Probability of Referral Following Job Separation	π_0	0.228	0.00045
Probability of Referral Following Job Arrival	π_1	0.794	0.00073
Probability of Referring an Unemployed Friend	ν_u	0.626	0.00051
Probability of Referring a School Friend	ω^{Edu}	0.155	0.00013
Probability of Referring a Work Friend	ω^{Work}	0.653	0.00014

Table 1.8: Parameter Estimates

sources of job dynamics through simulation and report the results in Table 1.9. The table shows that for non-executives' UE transitions, 72.15% result from direct job offers, 1.60% from referrals following friends' job arrivals, and 26.26% from referrals following job separations. For executives' EE transitions, the proportions are 34.00%, 3.07%, and 62.93% respectively; and for their wage raises, the proportions are 17.91%, 5.37%, and 76.73% respectively. These numbers show the following patterns. First, for all three types of transitions, referrals following friends' job separations lead to far more transitions than those following friends' job arrivals. This is due to the large number of job separations compared to the number of on-the-job offer arrivals, even though the probability of sending a referral is lower in the former case than the latter. Second, referrals contribute to higher proportions of job dynamics for executives (EE transition and wage raise) than for non-executives (UE transition). This suggests that the ratio of referral arrival rate and direct job offer arrival rate is larger for the executives than the non-executives. Third, for executives, referrals contribute to a higher fraction of wage raises than job-to-job transitions. The reason is that compared with direct offers, referred jobs are less likely to be more productive than executives' current jobs as they are no better than friends' own jobs.

	Panel A: Source of Accepted Offer in UE Transition		
	Direct Offer	Referrals Following Friends' Job Arrivals	Referrals Following Friends' Job Separations
Number	1851.7	41.0	673.9
%	72.15	1.60	26.26
	Panel B: Source of Accepted Offer in EE Transition		
	Direct Offer	Referrals Following Friends' Job Arrivals	Referrals Following Friends' Job Separations
Number	37.6	3.4	69.6
%	34.00	3.07	62.93
	Panel C: Source of Competing Offer in Wage Raise		
	Direct Offer	Referrals Following Friends' Job Arrivals	Referrals Following Friends' Job Separations
Number	34.7	10.4	148.7
%	17.91	5.37	76.73

Table 1.9: Sources of Job Dynamics: Direct Offers and Referrals

Notes: The reported numbers are averages over 100 simulations.

Model Fit

Figure 1.2 shows the fit for executive compensation conditional on firm productivity and number of executive friends. The model is able to replicate the empirical pattern that executive compensation is increasing in both the firm productivity and the number of executive friends. Figure 1.3 shows the fit for the productivity distribution, in which

I compare the empirical distribution of the socially isolated individuals' first jobs in the data and the parametrically estimated probability density function in the model. The model is capable of capturing the general shape of the empirical histogram. Table 1.10 shows the fit for the moments in job transitions. The model is able to replicate the spell durations and the frequencies of transitions in the model. Additionally, it is able to replicate the co-movement pattern that the probability of career advancement is higher when some friends leave current jobs than when no friend does. I show the fit for co-movement moments for each type of social connection separately in an additional Table 3.3 in Appendix 3.4.

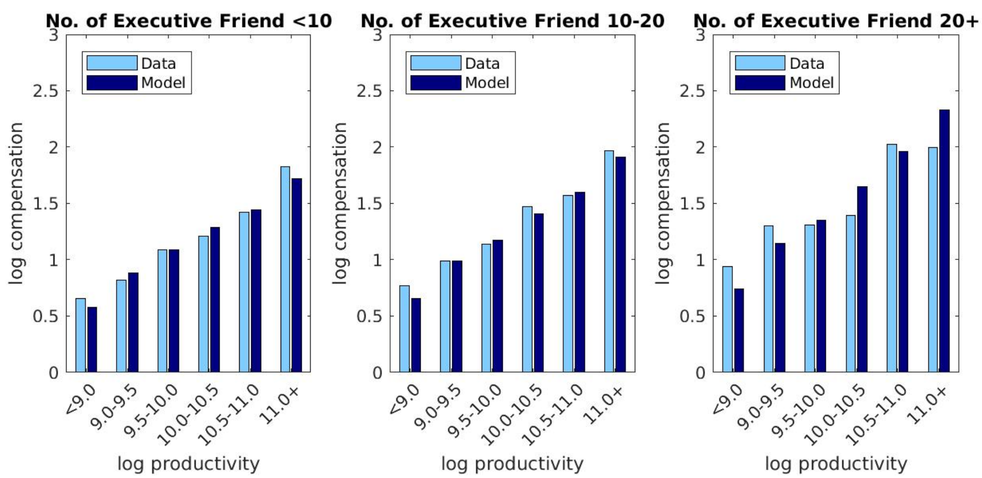

Figure 1.2: Model Fit for Executive Compensation

Notes: (1). The empirical measurement for firm productivity is market capitalization. (2). The unit for compensation and market capitalization are both 2015 U.S. million dollar.

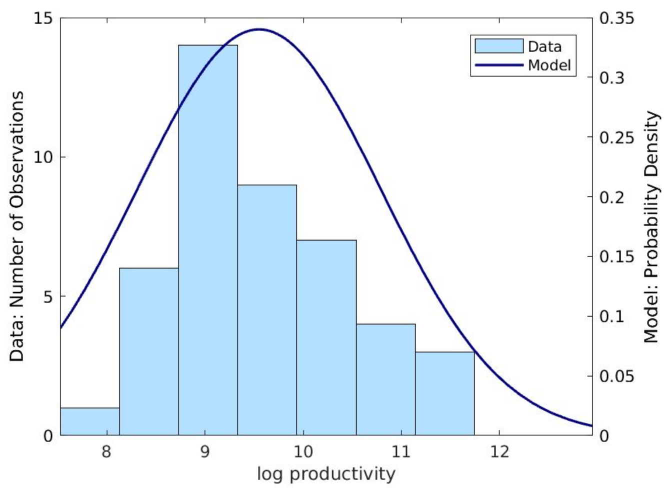

Figure 1.3: Model Fit for Productivity Distribution

Notes: (1). The empirical measurement for firm productivity is market capitalization, and the unit is 2015 U.S. million dollar. (2). The histogram plots the distribution of the socially isolated individuals' first jobs' productivities.

Moments	Data	Model
Mean of Employment Spell	3.9193	4.1457
Std. Deviation of Employment Spell	2.3548	2.8232
Mean of Unemployment Spell	4.4119	4.0953
Std. Deviation of Unemployment Spell	2.8509	2.8276
Frequency of UE Transition	0.6076	0.6109
Frequency of EE Transition	0.5778	0.5832
Frequency of EU Transition	0.0303	0.0296
Fraction of unemployed individuals experiencing UE transition		
...... if there is no friend experiencing EU transition	0.0812	0.1344
...... if there are friends experiencing EU transition	0.1717	0.1753
...... if there is no friend experiencing EE transition	0.1541	0.1603
...... if there are friends experiencing EE transition	0.1870	0.1926
Fraction of employed individuals experiencing EE transition		
...... if there is no friend experiencing EU transition	0.0028	0.0031
...... if there are friends experiencing EU transition	0.0077	0.0078
...... if there is no friend experiencing EE transition	0.0054	0.0053
...... if there are friends experiencing EE transition	0.0130	0.0156

Table 1.10: Model Fit: Moments in Job Transitions

Notes: (1). A UE transition is defined as the transition from non-executive to executive. An EE transition is defined as the transition from one executive job to another with a productivity increase. An EU transition is defined as the transition from executive to non-executive. (2). Frequency of transition is calculated as the total number of transitions during sample period divided by the number of individuals. (3). The table reports the co-movement moments (last 8 rows) for the simplified network, which does not distinguish between different types of social connections.

1.6 Counterfactual Experiments

In this section, I use the estimated model to perform two sets of counterfactual experiments. First, I evaluate the effect of referrals by varying the referral probabilities. Second, I examine the effect of network structure by comparing the observed networks with randomly formed networks.

Varying Referral Probabilities

In this set of counterfactual experiments, I vary the referral probabilities, π_0 and π_1. In the baseline model, these probabilities are estimated to be $\pi_0 = 0.228$ for referrals following friends' job separations and $\pi_1 = 0.794$ for referrals following friends' job arrivals. In two counterfactual experiments, I set the referral probabilities to be zero ($\pi_0 = 0,\ \pi_1 = 0$) and one ($\pi_1 = 1,\ \pi_1 = 1$) respectively.

I compare the outcomes in the counterfactual models with the baseline model, in terms of the number of referrals received and the executive's lifetime utility measured by consumption equivalence. In Table 1.11, I break down the individuals by their network degrees and report their outcomes. As expected, for both counterfactual experiments and both outcome measures, referrals have stronger effects for higher degree individuals. High degree individuals have higher gains from referrals partly because they receive more referrals, and partly because they can extract a higher share of the surplus in wage bargaining once they receive a referral.[43]

In the case of setting referral probabilities to zero ("no referral"), the lowest degree group loses 0.17 referrals (in a total time span of 9 years), which is equivalent to a

[43]Recall that bargaining power depends on the number of executive friends. This heterogeneity generates, even in the absence referrals ($\pi_0 = \pi_1 = 0$), a difference in individual welfare (consumption equivalence) across degree groups.

2.62% reduction in annual income; the highest degree group loses 0.35 referrals, which is equivalent to a 6.93% reduction in annual income.

In the other case of setting referral probabilities to one ("mandatory referral"), the changes in outcomes are more significant. For all degree groups, the change generates more than 0.5 additional referrals, which is equivalent to a 6.73% increase in annual income for the lowest degree group and a 15.96% increase for the highest degree group.

Degree	$d < 20$	$20 \leq d < 40$	$40 \leq d < 60$	$60 \leq d < 80$	$d >= 80$
	Number of Referrals Received				
Baseline	0.17	0.22	0.25	0.31	0.35
$\pi_0 = 0, \pi_1 = 0$	0.00	0.00	0.00	0.00	0.00
Δ	-0.17	-0.22	-0.25	-0.31	-0.35
$\pi_0 = 1, \pi_1 = 1$	0.56	0.71	0.80	0.99	1.07
Δ	0.39	0.49	0.55	0.68	0.71
	Annual Consumption Equivalence (million \$)				
Baseline	1.35	1.43	1.53	1.62	1.82
$\pi_0 = 0, \pi_1 = 0$	1.32	1.37	1.46	1.51	1.68
$\%\Delta$	-2.62%	-3.59%	-4.26%	-6.31%	-6.93%
$\pi_0 = 1, \pi_1 = 1$	1.45	1.56	1.72	1.86	2.12
$\%\Delta$	6.73%	9.38%	12.19%	15.09%	15.96%
Group Size	1,684	1,188	593	289	438
mean(Degree)	11.58	28.30	48.40	68.62	145.60

Table 1.11: Counterfactual Experiment: Varying Referral Probabilities

Notes: (1). All the changes are calculated with respect to the baseline model, in which referral probabilities are $\pi_0 = 0.228$ and $\pi_1 = 0.794$. (2). The reported numbers are averages over 100 simulations. (3). The time span used in the simulation is 9 years, the same as the sample length. (4). d is the degree for $Y^{All,2015}$, the simplified network in the final sample period.

Random Networks

In this counterfactual experiment, I vary the network structure. Specifically, I generate a new set of education, work, and social-activity networks such that the degree of each individual in each year is the same as the observed networks (baseline), but the connections are formed at random.[44] I find that the welfare distribution is more unequal under the random networks, and I investigate the underlying mechanism through the lens of two local network statistics.

Differences in the Local Network Structures

Before I present the results, it is useful to first examine the difference between the structure of the observed networks and the random networks. By design, the two sets of networks have equal degrees for all individuals at all time. The main difference lies in two other local structures: friends' popularity and local clustering.

Friends' Popularity. I measure friends' popularity by average friend degree. Define i's average friend degree in network k to be

$$\frac{\sum_{j \in N^k(i)} {d^k}_j}{|N^k(i)|}, \tag{1.30}$$

the average degree of i's friends'. It can be interpreted as the level of "popularity" of friends.

The patterns of the difference in average friend degree between the two sets of networks are heterogeneous with respect to own degree. As described in Section 1.3, the observed networks exhibit sorting on degree. Low-degree individuals tend to

[44]I describe the algorithm for simulating random network with a given degree sequence in Appendix 3.5.

have more low-degree friends in the observed network. Therefore, they have lower average friend degree in the observed network, compared with the random network. High-degree individuals tend to have more high-degree friends in the observed network. Therefore, they have higher average friend degree in the observed network, compared with the random network. Figure 1.4 breaks down the individuals by their own degrees and plots the distributions of average friend degrees for the observed and the random networks respectively.

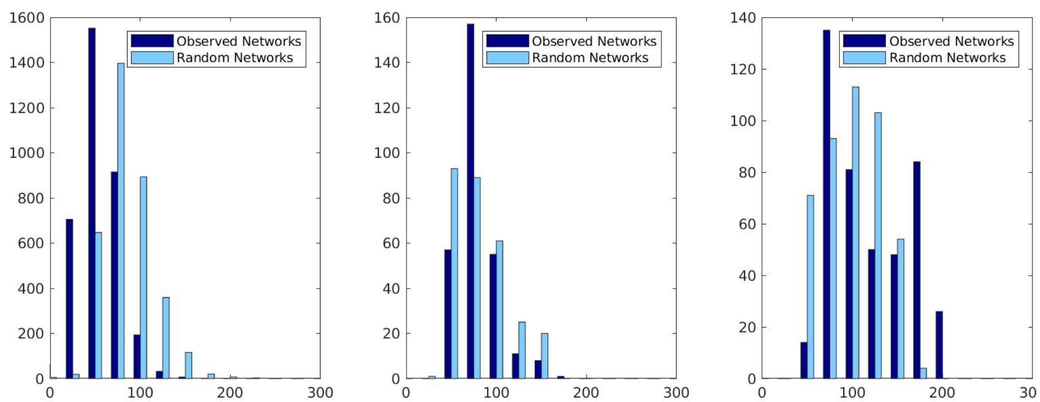

Figure 1.4: Distribution of Average Friend Degree for Low (Left), Medium (Middle), High (Right) Degree Individuals

Notes: (1). Average friend degree, defined in Eq.(1.30), gives the average degree of an individual's friends. (2). The figure on the left uses the subset of individuals with degree less than 60; the figure in the middle uses the subset of individuals with degree greater than or equal to 60 and less than 80; and the figure on the right uses the subset of individuals with degree greater than or equal to 80. (3). The degree and the average friend degree are calculated based on the simplified network in the final sample period, $Y^{All,2015}$, for the observed and the random network respectively.

Local Clustering. I measure the level of clustering by the local clustering coefficient defined in (1.3). It gives the fraction of an individual's friends who are also friends with one another.

The observed networks exhibit universally a far higher level of clustering than the random networks. Figure 1.5 plots the distributions of local clustering coefficients for the observed and the random networks respectively. In the observed networks, 88% of the individuals have clustering coefficients above 0.2, and 50% above 0.35. In stark contrast, in the random networks, 97.50% of the individuals have clustering coefficients below 0.2, and 50% below 0.03.

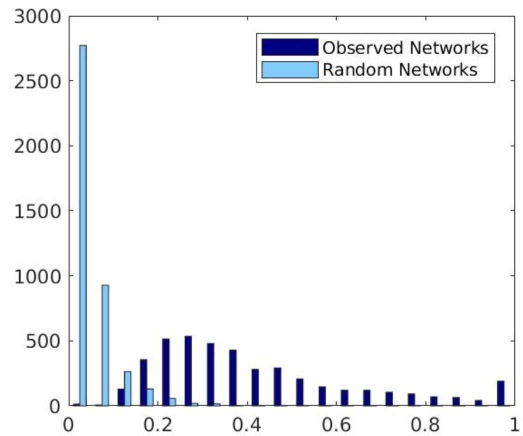

Figure 1.5: Distribution of Clustering Coefficient

Notes: (1). Clustering coefficient, defined in Eq.(1.3), gives the fraction of an individual's friends who are also friends with one another. (2). The clustering coefficients are calculated based on the simplified network in the final sample period, $Y^{All,2015}$, for the observed and the random network respectively.

Effects of Network Structures on Referrals

In this part, I consider a change from the observed networks to random networks with the same degree sequences and examine the resulting changes in referral patterns as well as executive welfare.

Overall Effects. Figure 1.6 shows that the random networks lead to greater inequality, compared with the observed networks. Under random networks, there are more individuals with low welfare and also more individuals with high welfare, which is a result of increased inequality in referrals. Table 1.12 breaks down the individuals by degree and reports the changes in their local network structures and labor market outcomes. The decrease in the clustering coefficient is significant for all degree groups. The average friend degree increases for the lower degree groups and decreases for the highest degree group. In terms of the labor market outcomes, low-degree individuals are worse off under the random networks, whereas high-degree individuals are better off, as shown in their welfare measured by consumption equivalence. For example, the lowest degree group receives 0.10 fewer referrals, which is equivalent to a 1.41% reduction in annual income. The highest degree group receives 0.20 more referrals, which is equivalent to a 3.52% increase in annual income.

Effects of Friend Popularity and Clustering by Degree Group. I investigate the mechanisms for the overall effects through the lens of two local network structures discussed in Section 1.6, friends' popularity and local clustering. As discussed in Section 1.2, the qualitative effects of these network structures on job referrals are ambiguous. First, friends' popularity has two countervailing effects on the arrivals of referrals. On the one hand, a popular friend means high competition for referrals, lowering a worker's probability of receiving a particular referral sent by his/her popular friend (competition effect). On the other hand, a popular friend benefits from his/her large set of friends' referrals, increasing the quantity and the quality of referrals he/she sends out (ripple effect). Second, local clustering also has two countervailing effects on the arrivals of referrals. An advantage of high clustering is that it keeps the positive

Degree Quintile	1	2	3	4	5
	Clustering Coefficient				
Observed Networks	0.70	0.49	0.35	0.31	0.31
Random Networks	0.04	0.04	0.04	0.05	0.07
Δ	-0.65	-0.45	-0.30	-0.26	-0.23
	Average Friend Degree				
Observed Networks	35.74	48.52	58.84	68.67	98.80
Random Networks	89.91	84.40	83.23	84.05	91.14
Δ	54.17	35.88	24.40	15.39	-7.66
	Number of Referrals Received				
Observed Networks	0.16	0.18	0.21	0.25	0.33
Random Networks	0.06	0.10	0.16	0.27	0.52
Δ	-0.10	-0.08	-0.05	0.02	0.20
	Annual Consumption Equivalence (million $)				
Observed Networks	1.34	1.37	1.42	1.49	1.71
Random Networks	1.32	1.36	1.42	1.49	1.77
%Δ	-1.41%	-0.51%	0.04%	0.39%	3.52%
Group Size	719	860	898	854	861
mean(Degree)	7.32	14.23	24.26	40.94	105.89

Table 1.12: Counterfactual Experiment: Random Networks

Notes: (1). All the changes are calculated with respect to the observed networks. (2). The reported labor market outcomes are averages over 100 simulations. (3). The time span used in the simulation is 9 years, the same as the sample length. (4). Degree, average friend degree, and clustering coefficient are defined in (1.2), (1.30), and (1.3) respectively. They are calculated based on the simplified network in the final sample period, $Y^{All,2015}$, for the observed and the random network respectively.

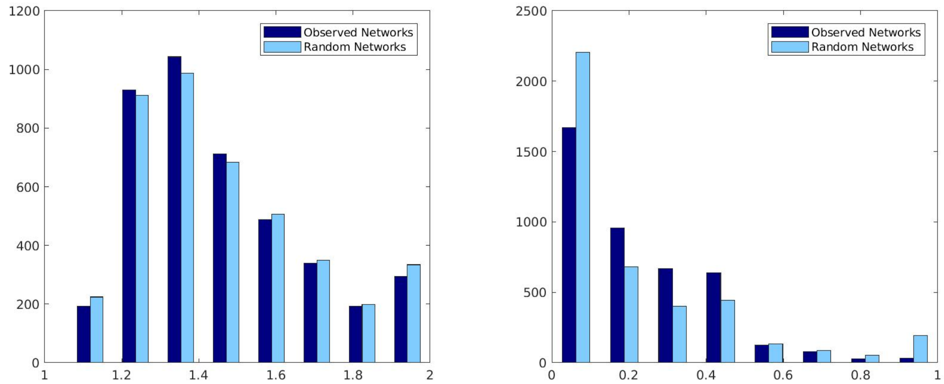

Figure 1.6: Distributions of Executive Welfare (Left) and Number of Referrals Received (Right)

Notes: (1). Executive welfare is measured by annual consumption equivalence in million \$. (2). In the plot of distribution of executive welfare, all individuals with consumption equivalence higher than 2 million are plotted at 2 million. (3). In the plot of distribution of the number of referrals received, all individuals with more than 1 referral are plotted at 1 referral. (4). The numbers are averages over 100 simulations. (5). The time span used in the simulation is 9 years, the same as the sample length.

spillovers in an inner circle (closeness effect). A disadvantage is that it limits the positive spillovers from a distance (isolation effect).

I quantitatively analyze these network effects by regressing the percentage change in the number of referrals an individual receives on the change in his/her average friend degree and the change in his/her local clustering coefficient. I allow heterogeneous effects with respect to an individual's own degree by interacting the regressors with degree group dummies. Table 1.13 reports the results.

The results show that for all degree groups, the marginal impact of higher friend popularity is negative. This is driven by the competition effect from friends of friends. As the change from the observed networks to the random networks induces heterogeneous changes in friend popularity for different degree groups, the welfare implication will also be heterogeneous.

Dependent variable: Δ% Number of Referrals			
(ΔAvg. Friend degree)	$\times \mathbb{1}(\text{degree} < 20)$	-0.0056***	(0.0008)
...	$\times \mathbb{1}(20 \leq \text{degree} < 40)$	-0.0081***	(0.0011)
...	$\times \mathbb{1}(40 \leq \text{degree} < 60)$	-0.0074***	(0.0016)
...	$\times \mathbb{1}(60 \leq \text{degree} < 80)$	-0.0085**	(0.0028)
...	$\times \mathbb{1}(80 \leq \text{degree} < 100)$	-0.0149**	(0.0047)
...	$\times \mathbb{1}(100 \leq \text{degree} < 150)$	-0.0095	(0.0056)
...	$\times \mathbb{1}(\text{degree} \geq 150)$	-0.0189*	(0.0074)
(ΔClustering coef.)	$\times \mathbb{1}(\text{degree} < 20)$	-0.0009	(0.1127)
...	$\times \mathbb{1}(20 \leq \text{degree} < 40)$	-0.6455***	(0.1783)
...	$\times \mathbb{1}(40 \leq \text{degree} < 60)$	-0.8782***	(0.2514)
...	$\times \mathbb{1}(60 \leq \text{degree} < 80)$	-0.4454	(0.4031)
...	$\times \mathbb{1}(80 \leq \text{degree} < 100)$	0.4937	(0.6450)
...	$\times \mathbb{1}(100 \leq \text{degree} < 150)$	1.5269	(0.9156)
...	$\times \mathbb{1}(\text{degree} \geq 150)$	4.3564***	(1.2196)
Observations		4,192	

Table 1.13: Effects of Friend Popularity and Clustering on Job Referral

Notes: (1). This table summarizes the results of an OLS regression, where the dependent variable is the percentage change in the number of referrals received. (2). Degree, average friend degree, and clustering coefficient are defined in (1.2), (1.30), and (1.3) respectively. They are calculated based on the simplified network in the final sample period, $Y^{All,2015}$, for the observed and the random network respectively. (3). Other covariates are degree, degree^2, degree^3. (4). Standard errors are reported in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Moreover, the marginal impact of higher clustering is negative for low-degree groups, but positive for high-degree groups. The intuition for the heterogeneous effects of clustering is the following. When an individual's degree is low, the number of good shocks among his/her immediate friends is low, so it is relatively more important to be able to benefit from the spillovers from a distance. High clustering limits such chances, generating a negative impact on referrals (isolation effect dominates). When an individual's degree is high, however, the number of good shocks among his/her immediate friends is large enough, so it is relatively more important to keep these spillovers in an inner circle. High clustering provides such protections, generating a positive impact on referrals (closeness effect dominates). As the change from the observed networks to the random networks decreases the level of clustering universally, the heterogeneous effects of clustering for different degree groups generate heterogeneous welfare implication.

Qualitative Discussion. In summary, I present the change in two network statistics in Section 1.6, the marginal effects of these changes in Table 1.13, and the overall effects on referrals and thus welfare in Table 1.12. Many of these changes and effects are heterogeneous, so it is useful to provide a qualitative summary. Table 1.14 summarizes the qualitative effects of the changes in network structure on referrals through the lenses of the two local network structures discussed above.

	Low Degree	High Degree
Friend Popularity		
Δ Avg. Friend Degree	increase	decrease
Marginal Effect	negative	negative
Implied Change in Referrals	decrease	increase
Clustering		
Δ Clustering Coefficient	decrease	decrease
Marginal effect	negative	positive
Implied Change in Referrals	increase	decrease
Overall		
Change in Referrals	decrease	increase

Table 1.14: Qualitative Effects of a Change from the Observed Networks to the Random Networks

For low-degree individuals, their friend popularity increases, and a negative marginal effect implies a decrease in referrals. Their local clustering decreases, and a negative marginal effect implies an increase in referrals. Overall, the number of referrals–and thus welfare–decreases.

For high-degree individuals, their friend popularity decreases, and a negative marginal effect implies an increase in referrals. Their local clustering decreases, and a positive marginal effect implies a decrease in referrals. Overall, the number of referrals–and thus welfare–increases.

This counterfactual experiment highlights the importance of the network structure beyond the number of friends. Although the observed and the counterfactual networks have the same degree sequences, their difference in connection patterns leads to different welfare distribution. Specifically, with high-degree individuals gaining and low-degree individuals losing under the random networks, the distribution of referrals, and thus welfare, is more unequal. This counterintuitive finding further demonstrates the delicacy of the mechanisms in network models and the need for rigorous study.

1.7 Conclusion

In this paper, I study the network effects of job referrals and investigate the impact of social network structure on workers' labor market outcomes. I develop a job search model that incorporates referral and non-referral job offers and different kinds of social networks. In my model, the network structure is dynamic and evolves as workers move across jobs. Workers directly receive job offers as well as referrals from their friends. Referrals are generated endogenously: an external referral occurs when a friend rejects an offer that he/she receives, and an internal referral occurs when a friend leaves his/her current job. As a result of this referral process, the quantity and the quality of referrals depend not only on the worker's number of friends but also on the quality of these friends' jobs. Moreover, the model incorporates the full network structure beyond the number of friends, and it generates rich network effects beyond immediate friends.

My empirical analysis combines three different data sources on the corporate executive labor market. I first provide reduced-form evidence on job referrals in the executive labor market. I then estimate the structural job search model by Generalized Method of Moments (GMM). The estimation results show both the statistical and economical significance of job referrals. My model nests a model with no referrals, and a specification test rejects such a model. Simulations from my model show that more than one quarter of job transitions and raises are driven by referrals.

I use the estimated model to perform two counterfactual experiments. The first experiment evaluates the welfare effect of referrals by varying the probability of referrals. I find that shutting down referrals reduces executives' welfare by an equivalence of a 2-7% decrease in annual income and that mandatory referral boosts executives' welfare by an equivalence of a 6-16% increase in annual income.

The second experiment examines the welfare effect of network structure by varying the network structure. Specifically, a new set of counterfactual networks are generated in which the individuals have the same number of friends as the observed networks, but the connections are formed randomly. I find that the welfare distribution is more unequal under the randomly formed networks. I further investigate the mechanisms for these effects through the lens of two network statistics: friends' popularity and local clustering, and my findings are as follows. First, in terms of friends' popularity, the competition effect dominates the ripple effect. Second, in terms of local clustering, the isolation effect dominates for individuals with a small number of friends, and the closeness effect dominates for individuals with a large number of friends. Overall, the competition effect resulting from the change in friend popularity dominates, generating greater inequality under the random networks. This experiment highlights the effects of network structure beyond the number of friends.

This paper focuses on worker heterogeneity in network position. An interesting extension is to incorporate more heterogeneity such as gender into the analysis and study how referrals and social networks affect the gender pay gap and the underrepresentation of female executives. Additionally, this paper abstracts from the roles played by intermediaries such as executive search firms in soliciting referrals as well as the executives' strategic concerns in providing referrals. An interesting extension to the model is to explicitly incorporate the executive search companies and study their interactions with the executives in generating referrals. These are exciting areas for future research.

Chapter 2

Inferring the Ideological Affiliations of Political Committees via Financial Contributions Networks

This chapter is co-authored with Hanming Fang.

2.1 Introduction

Campaign finance is an integral part of the U.S. politics, and political committees are major participants in campaign finance related activities. We use the term *Political Committees* (henceforth, PCs) to refer to federal political action committees (PACs), party committees, campaign committees for presidential, house and senate candidates, as well as issue-based groups or organizations, including lobbyists or fundraisers. They collect contributions from individual donors, make contributions to other committees and candidates, and spend money for or against candidates.

U.S. campaign finance laws mandate that political committees disclose all financial transactions including the contributions they receive and their expenditures, thus rendering numerous data for analyzing their campaign related activities. However, the PCs are *not* mandated to file their party (or ideological) affiliations. Indeed, as we will detail in Section 2.3, nearly 60% of PCs' party affiliations are unreported. This missing data problem may generate obstacles in the study of important issues related to campaign finance. For example, for researchers who want to study the patterns of individual contributions using the individual contributions data provided by the Federal Election Commission (FEC), it is important to know the ideological affiliations of the PCs to which individual donors contribute. Nevertheless, in the Contributions by Individuals (Year 2003-2004) data released by the Federal Election Commission (FEC), about 24%, both in terms of the instance and the amount, of individual contributions go to PCs with unknown party affiliations.

In this paper, we contribute to the methodologies that aim to address the missing ideological affiliations of the PCs. We propose a method of inferring PCs' ideological affiliations from the financial transactions network among the PCs. The Contributions to Committees from Committees data, also administered by the FEC, contains the universe of all records of financial contributions among all registered PCs. We use this data set of the contribution activities to construct a financial transactions network of the PCs, where each PC is a vertex of the network, and the money flows between the PCs form the (weighted) edges. The basic idea of our method of inferring ideological affiliations of the PCs is simple. If the PCs tend to contribute more frequently to other PCs with similar ideology, then the PCs that actually filed their party affiliations, together with the structure of the observed financial transactions network, should provide information about the ideologies of the PCs with unknown ideological affiliations. Specifically, we build an economic model of link formation and transfer amount, and use the ideas of "community detection" first developed in the *stochastic block* model literature where contribution decisions (both the link formation and the weights) depend on both the observed characteristics and the *potentially* latent (for the PCs that do not file their party affiliations) ideologies of the potential contributing and receiving PCs.

Our model incorporates several new features that are absent in the existing stochastic block model literature. First, we introduce weights and rich heterogeneity in the edge formation process: the decision to make a contribution and its contribution amount is not only governed by the latent political ideologies of the PCs, but also depends on the vertex-level contextual information such as financial and institutional characteristics. Second, we model the reported party affiliations of those PCs that do self report their party affiliations as noisy measurements of their true latent ideologies. Thus, our methodology allows us to estimate the latent ideologies of all PCs, including those that self reported party affiliations.

We use three publicly available data sets in our analysis. Two are from the campaign finance record in 2003-2004 election cycle, namely the Contributions to Committees from Committees, and the Committee Master File, both maintained by the FEC. We use the first data set to construct the financial transactions network of the PCs, and we use the second data set to obtain the party affiliations of some PCs (if they self report), as well as the designations and types of all PCs. The third data set is additional industrial breakdown information of the PCs which we collected from `OpenSecret.org`.[45] Our data sets cover the *universe* of all PCs engaging in transactions with other PCs in 2003-2004. This feature of our network data has both advantages and disadvantages. An advantage is that it avoids potential bias arising from analyzing a partially sampled network, but a major disadvantage is the computational burden associated with the large network.[46] There are 5,858 vertices (i.e., PCs) in the financial transactions network, 3,727 of which did not report their party affiliations. Thus the number of potential ideological configurations is

[45] `https://www.opensecrets.org/pacs/list.php`

[46] Chandrasekhar and Lewis (2011) shows that bias arises when one works with *partially* sampled network.

enormous, specifically, 2^{3727} even if we just allow for the binary ideology of Democrat or Republican. For similar reasons, in a Bayesian approach that delivers a probability distribution over different ideologies for each vertex (see Section 2.4 for details), exact estimation of the posterior mode is infeasible. We instead propose a Gibbs sampler algorithm to approximate the solution. In Monte Carlo Simulations, we demonstrate that our estimation algorithm achieves very high accuracy in recovering the latent party affiliations provided that the pairwise difference in ideology groups' connection patterns satisfy what is known as the Chernoff-Hellinger divergence criterion. We illustrate our approach using the campaign finance record in the 2003-2004 election cycle. Using the posterior mode to categorize the party affiliations of the PCs, our estimated ideological affiliations match the self reported ideology for 94.36% of those committees who self reported to be Democratic and 89.49% of those committees who self reported to be Republican.

Related Literature. This paper is closely related to two strands of existing literature. In terms of the research question, our paper is related to the political economy literature on the measurement of political ideologies. In terms of methodology, it is related to the statistics literature on community detection.

We first discuss the literature on the measurement of political ideology. In a seminal paper, Poole and Rosenthal (1985) proposed a measure of the ideology points (NOMINATE score) of federal legislators using the roll call data. In their paper, legislators' ideology points can differ by election cycle, and bridge legislators and bridge bills are used to ensure that the measures are comparable across time.[47]

[47] "Bridge legislators" are legislators who serve multiple terms; and "bridge bills" are bills that are considered at different legislative cycles but with similar contents.

Note that Poole and Rosenthal's NOMINATE score is only available for members of Congress, which is a very small sample (around 500 legislators in each election cycle); and it is tied to voting behavior. Subsequently, ideology measurement for other political actors are proposed based on the NOMINATE score. For example, McCarty et al. (2006) combined NOMINATE score and the amount of contribution from PACs to the legislators to estimate PACs' ideology. Their proposed measure is money-weighted average of the NOMINATE scores of the legislators to whom the PAC has contributed. These measures could be biased because they do not account for PACs' contribution to losing candidates and other political actors. McKay (2008) and McKay (2010) combined NOMINATE score and the preferred votes on key roll calls published by interest groups to estimate interest groups' ideologies. Her proposed measure is the average of the NOMINATE scores of "perfectly scoring" legislators whose votes are exactly the same as the preferred votes of the particular interest group. These measures are only available for interest groups which publish their preferred votes. Additionally, they could be biased: if an interest group publishes many key votes, the number of "perfectly scoring" legislators could be too small and leads to inaccuracy; if an interest group publishes only a few key votes, the number of "perfectly scoring" legislators could be too large and artificially draws the measure toward the center.

There are other proposed methods which do not rely on the NOMINATE score. Some studies use the campaign finance data to jointly estimate candidates and PACs' ideologies. For example, McCarty and Poole (1998) proposed a measure based on PAC's contribution decision between incumbent-challenger pairs, excluding unchallenged and open seat elections which account for a significant fraction in the federal elections. Their measures are not available for candidates in these elections; and are potentially biased for PACs which contributed in these elections. More recently, Bonica (2013) proposed a method using the contributions from PACs to candidates. This approach, from the perspective of network study, restricted the sample to a directed bipartite graph, excluding connections among PACs or candidates, as well as connections from candidates to PACs. In our paper, in contrast, we use an unrestricted network incorporating all financial connections. Moreover, he uses maximum likelihood estimation method which requires multiple observations, so he pooled observations over a period of 30 years (1980-2010) and further restricted the sample to PACs that have given to 30 or more unique candidates, and candidates who have received from 30 or more unique PACs. Pooling data over time is potentially problematic. It requires the assumption that the actor's ideology is fixed in a span of 30 years. Differently, in our paper, we make inference about a PC's ideology out of a single observation of financial transactions network: for any PC, separate ideology measures can be calculated for each election cycles. As a result, our measures are well suited for the study of the time trend of political activities. Moreover, his sample selection excludes candidates and PACs with small numbers of financial transactions. However, in our paper, we cover a more extensive scope of political actors: a PC is included as long as it has at least one financial transaction with another PC.

Other studies use data from the social media platforms. For example, Barberá (2015) used the Twitter "following" links to estimate the ideology of the political elites (accounts of candidates, parties, media, and journalists) and the mass (other individual accounts) in multiple countries on the same scale. Similar to Bonica (2013), he restricted the sample to a directed bipartite graph focusing on the mass following the elites. He used a Bayesian method and a two-stage estimation strategy that exploits the bipartite structure. In the first stage, he used a sub-network induced by a random subsample of the mass accounts and all the elite accounts, and jointly estimated their ideologies. In the second stage, holding the first stage ideology measures of the elite accounts fixed, he estimated the ideology for all the mass accounts using the full network. This two stage procedure reduces the computational intensity; but using a partially sampled network may lead to bias in the first stage, which may be further propagated in the second stage. Our approach is different in that we simultaneously estimate the ideologies of all PCs using the full network. Finally, one key difference between Bonica (2013) and Barberá (2015) and our paper is that both of these papers use spatial models. Spatial models assume *a priori* homophily: political actors with closer ideology points have higher propensity to connect. This has an especially strong implication on the centrists' behavior. It rules out the possibility for the centrists to have low propensity to connect with other centrists, and high propensity to connect with the center-left and the center-right. In contrast, our model can accommodate non-homophilic patterns. In fact, according to our estimates, the Independent PCs do have higher propensity to form financial connections with Democratic and Republican PCs than other Independent PCs.

We now discuss the literature of community detection. The main task in this literature is to classify vertices in a network into different groups based on observed connections. When repeated observations of the network are available, canonical statistical tools can be directly applied. For example, Trebbi and Weese (2015) used generalized method of moments to estimate, for each district (vertex), the fraction of insurgents in different groups. The number of parameters is of order n, the number of vertices. Though large, it is fixed when the number of network observations T grows. Therefore, standard asymptotic results and inference tools are valid.

When only one observation of the network is available, which is the case for our application, the nature of the problem is changed. In this case, the only hope for asymptotic result is to have network size $n \to \infty$; however, when the network size grows, the number of parameters (of order n) also grows, which renders canonical statistical tools invalid. Some popular approaches circumvent this issue by using model-free heuristics: minimum-cut method in Stoer and Wagner (1997) minimizes the number of edges between communities; modularity maximization method in Newman (2006) maximizes the difference between the fraction of the edges within groups and its expected value if edges were formed at random; convergence of iterated correlations (CONCOR) method in Breiger et al. (1975) bisects the adjacency matrix by iteratively calculating correlation coefficients among rows (or columns); and spectral method in Newman (2013) extracts information of graph partition from the top few eigenvectors of the adjacency matrix. Though widely used in practice, these approaches have the following limitations. First of all, the first three methods all assume a priori assortative communities, i.e., denser within-community connection than between-community connection. They are not appropriate for problems where some group may have higher external connectivity. For example, in our application, the Independent PCs may engage in more transactions with Democratic and Republican PCs than with other Independent PCs. We need a model which does not assume away this possibility. Second, in the absence of a statistical model, none of these methods allows for statistical inference: we cannot state how confident we are in the obtained classification. Third, these methods cannot incorporate vertex level or vertex pair level heterogeneity beyond the latent community.

Therefore, we took a model-based approach instead - we build on the stochastic block model (SBM) initiated by Snijders and Nowicki (1997) and Nowicki and Snijders (2001). In the SBM, a network is randomly formed conditional on the underlying community structure, and the community structure itself is also stochastically generated. This model is widely accepted as a canonical model for community detection and its estimators have desirable properties. Recent studies characterized the information-theoretic threshold for exact recovery, i.e. conditions on the data generating process such that one observation of the network embodies enough information to exactly recover the community structure. Mossel et al. (2014), Abbe et al. (2014), and Abbe and Sandon (2015) studied this problem in unweighted networks. Jog and Loh (2015) and Yun and Proutiere (2016) extended previous results to weighted networks. In particular, they show that if the Chernoff-Hellinger divergence of community pair's connection patterns is above a particular threshold, the probability of correctly recovering the entire community structure converges to 1 as $n \to \infty$. Kanade et al. (2016) and Cai et al. (2017) studied this problem when a fraction of the vertices's community affiliations are revealed. Finally, Abbe (2017) provided a detailed review of the recent development in the research of exact recovery in stochastic block models. As explained earlier, exact solution to the Bayesian posterior mode is infeasible for large network, and various algorithms are proposed to approximate the solution. For example, the spectral algorithm can be viewed as an approximation to posterior mode.[48] The Gibbs sampler algorithm we use in this paper is an approximation to first obtain the posterior distribution and then the posterior mode. From the perspective of

[48]The original discrete parameter space is relaxed to a sphere, and the spectral method is a solution to the relaxed problem, and is derived through derivatives. A detailed description can be found in Newman (2013).

empirical implementation, a weakness of the generic SBM is the lack of heterogeneity beyond community affiliations, so we introduce pair-wise heterogeneity in a similar way as Peng and Carvalho (2015a) and Peng and Carvalho (2015b). These papers proposed a degree correction strategy to capture vertex-level heterogeneity: they use additional latent variable or observed degree to capture the popularity of a vertex, and a "popular" vertex has higher propensities to connect to all other vertices. In this paper, we allow heterogeneity at *vertex pair* level, by incorporating interactions of the observable vertex level characteristics, such as location, industry, and budget. Apart from richer specification, the use of observable characteristics has additional benefits: it entails less computational intensity than the use of latent variable, and it avoids the endogeneity problem from the use of observed degrees.

The remainder of the paper is structured as follows. In Section 2.2, we introduce the basic framework for the financial network among political committees, and describe the statistical model of random network formation that we will empirically implement; in Section 2.3, we describe the data sets used in our analysis, as well as several "naive" methods that we attempted; in Section 2.4, we provide the details of the Bayesian estimation procedure and arguments for identification; in Section 2.5, we present the Monte Carlo results; in Section 2.6, we describe the specifications used in our empirical implementation, and present the main empirical results; finally, in Section 2.7, we conclude and discuss directions for future research.

2.2 Financial Network of Political Committees

We represent the money-flow network among political committees as a static, weighted and undirected graph.[49] A graph $\mathcal{G}$ consists of vertices $\mathcal{V}$ and edges $\mathcal{E}$. In our model, each vertex $i \in \mathcal{V} = \{1, ..., n\}$ represents a PC. In our application, n will represent the total number of PCs registered with the FEC, and is 5,858 for the 2003-2004 election cycle. Each edge $(i, j) \in \mathcal{E}$ is an unordered pair in $\mathcal{V} \times \mathcal{V}$. In our application, $(i, j) \in \mathcal{E}$ if there exists money flow, either unilateral or bilateral, between committees i and j. A weighted graph also includes, for each edge $(i, j) \in \mathcal{E}$, a corresponding weight w_{ij}, which in our application will correspond to the sum of money flows between the two PCs. Equivalently, a weighted graph $\mathcal{G}$ can be represented by a weighted adjacency matrix $\mathbf{y}$ where

$$y_{ij} = \begin{cases} w_{ij} & \text{if } (i, j) \in \mathcal{E} \\ 0 & \text{otherwise.} \end{cases} \tag{2.1}$$

In addition to the network structure described above, each committee $i \in \{1, ..., n\}$ is characterized by the following attributes: a unidimensional latent ideology x_i, and a multi-dimensional observable characteristics $\mathbf{z}_i$, which captures the financial and institutional characteristics of the PC. The details of the variables contained in the vector $\mathbf{z}_i$ will be described in Section 2.6 when we present our empirical specification.

Ideologies of Vertices. We assume that there are m discrete categories of ideologies where $m \geq 2$. In our application, m will be equal to 3, corresponding to Democratic, Republican and Independent respectively. We denote the *true* ideology of vertex $i \in \mathcal{V}$ by $x_i \in \{1, ..., m\}$. We assume that, for all vertices, their x_i's are latent and

[49]While both the model and the estimation strategy can be straightforwardly extended to a directed graph, an undirected graph is adopted for computational tractability.

unobserved to us. However, For a *subset* of vertices $\mathcal{V}^o \subset \mathcal{V}$, there exists a *noisy* measure, denoted by $\hat{x}_i \in \{1, ..., m\}$, of the latent ideology x_i; but for vertices in $\mathcal{V} \backslash \mathcal{V}^o$, we do not have this noisy measure. In our application, the size of $\mathcal{V}^o$ is typically about 1/3 of the size of $\mathcal{V}$.[50]

To summarize, the framework for the financial network of PCs can be represented by:

$$\langle \mathbf{y}, \mathbf{x}, \hat{\mathbf{x}}, \mathbf{z} \rangle = \langle \{y_{ij}\}_{1 \le i,j \le n},\ \{x_i\}_{1 \le i \le n},\ \{\hat{x}_i\}_{i \in \mathcal{V}^o},\ \{\mathbf{z}_i\}_{1 \le i \le n} \rangle . \tag{2.2}$$

Note, however, since we do not observe $\mathbf{x}$, the data consists of

$$\text{DATA} : \ \langle \mathbf{y}, \hat{\mathbf{x}}, \mathbf{z} \rangle = \langle \{y_{ij}\}_{1 \le i,j \le n}, \{\hat{x}_i\}_{i \in \mathcal{V}^o}, \{\mathbf{z}_i\}_{1 \le i \le n} \rangle . \tag{2.3}$$

The goal of our empirical exercise is to make *inference* about the latent ideology $\mathbf{x}$ based on data $\langle \mathbf{y}, \hat{\mathbf{x}}, \mathbf{z} \rangle$.

A Statistical Model

As we explained in the introduction, the vector of latent ideologies $\mathbf{x} = (x_1, ..., x_n)$ can be high-dimensional. To circumvent the high-dimensionality problem, we adopt a Bayesian approach by assuming that each vertex can, *a priori*, be of ideology $k \in \{1, ..., m\}$ with probability θ_k where $\boldsymbol{\theta} = (\theta_1, ..., \theta_m) \in \Delta^{m-1}$; and then we use its observed links with other vertices and the weights of the links to render a *posterior* probability distribution $\mathbf{p}_i \in \Delta^{m-1}$ over these categories. We will use the *mode* of the posterior distribution $\mathbf{p}_i$ as our best guess for i's political or ideological affiliation.[51]

[50] In Appendix 4.8, we also implement a version of the model in which we assume that $\hat{x}_i = x_i$ for all $i \in \mathcal{V}^o$.

[51] Of course, the whole posterior distribution $\mathbf{p}_i$ itself is of interest beyond its mode: $\mathbf{p}_i$ will allow us to provide a more continuous measure of $i's$ ideology spectrum. For example, two vertices may end up with the same ideological categorization based on their posterior mode, while one vertex could be more central than the other.

Formally, our model of the financial network formation among PCs is based on the *stochastic block* models of random network formation. In the model, the number of vertices n, the number of blocks m, and vertices' observable characteristics $\mathbf{z}$ are assumed to be exogenous and fixed. The adjacency matrix $\mathbf{Y}$, the ideology vector $\mathbf{X}$, and the noisy measure $\hat{\mathbf{X}}$ are assumed to be random.[52]

Prior. For any PC $i \in \mathcal{V}$, its latent ideology X_i has the following marginal prior distribution:

$$\mathbb{P}(X_i = k) = \theta_k \text{ for } k = 1, ..., m \tag{2.4}$$

where $\boldsymbol{\theta} = (\theta_1, ..., \theta_m) \in \Delta^{m-1}$.

Measurement Error. For a PC $i \in \mathcal{V}^o \subset \mathcal{V}$, we observed i's reported political affiliation $\hat{X}_i$, which we interpret as a noisy measure of its true ideology X_i. Specifically,

$$\mathbb{P}(\hat{X}_i = k | X_i = x_i) = \begin{cases} 1-\epsilon & \text{for } k = x_i \\ \frac{\epsilon}{m-1} & \text{for } k \in \{1, ..., m\} \backslash \{x_i\}, \end{cases} \tag{2.5}$$

where $\epsilon \in [0, 1)$ captures *the rate of measurement error.*[53]

Edge Formation. Conditional on the true ideologies, X_i, X_j, and observable characteristics $\mathbf{z}_i, \mathbf{z}_j$, entries Y_{ij}'s in the weighted adjacency matrix are assumed to be independently generated across (i, j) pairs:

[52]Throughout the paper, generic random variables are denoted by capital letters $\mathbf{Y}$, $\mathbf{X}$, and $\hat{\mathbf{X}}$, whereas their realizations and specific configurations are denoted by small letters $\mathbf{y}$, $\mathbf{x}$, and $\hat{\mathbf{x}}$.

[53]ϵ can be pre-specified as 0 if there is good reason to believe that there is no error in report (perhaps by the nature of the data set). Otherwise, a positive ϵ is more flexible because it allows for measurement error, which makes it possible for the report $\hat{x}_i$ to differ from the true ideology x_i. We will implement a version of the model restricting ϵ to be zero in Appendix 4.8, where we instead hold out a subset of the vertices with self-reported ideologies in our estimation, and use the holdout sample to validate our estimation results.

Assumption 1 (Conditional Independence). For any pairs of PCs, (i, j) and (i', j'), with either $i \neq i'$ or $j \neq j'$,

$$Y_{ij} \perp Y_{i'j'} | \left(X_i, X_j, \mathbf{z}_i, \mathbf{z}_j, X_{i'}, X_{j'}, \mathbf{z}_{i'}, \mathbf{z}_{j'} \right).$$

An economic interpretation of this assumption is as follows. A PC designs its general principle in contributions according to its political ideology as well as its financial and institutional characteristic. Following the principle, its staffs make decisions on whether to contribute to a specific committee, and the idiosyncratic factors in each of these decisions are unobserved by the researcher and are assumed to be independent. This conditional independence assumption is a potentially strong assumption, because it abstracts away from various possible strategic considerations that a political committee may have in its contribution decisions; of course, the restrictiveness of this assumption in practice depends on how complete the vector of characteristics $\mathbf{z}$ is. It is important to note that Assumption 1 is stated *conditional on the latent ideologies* of the relevant PCs. Thus, it allows for what we believe to be the first-order role of ideologies in political contributions. Other characteristics of the PC, for example, its previous connections, are also allowed to affect its contribution decisions, to the extent that such information is contained in its characteristics $\mathbf{z}$. In addition, it is well acknowledged that it is difficult to establish asymptotic results in a network model in the presence of widespread correlation (see, e.g., Leung (2016)); as a result, valid statistical inference using data from a single network (or a small number of networks) necessitates restrictions on the degree of correlation. Finally, conditional independence is a maintained assumption in the literature of stochastic block model (Snijders and Nowicki (1997) and Nowicki and Snijders (2001)), and the literature of classification in general (see, e.g., Koller and Friedman (2009)).[54]

Specifically, we assume that the edge formation process is as follows. For any pair of vertices $(i,j) \in \mathcal{V}^2$ with $i \neq j$, conditional on $(X_i, X_j) = (k,l) \in \{1,...,m\}^2$, and $\mathbf{z}_i, \mathbf{z}_j$,

$$\begin{aligned} Y_{ij} &> 0 \text{ if } \beta_{0,kl} + \boldsymbol{\beta_{1,kl}}(\mathbf{z}_i + \mathbf{z}_j) + \boldsymbol{\beta_{2,kl}}\mathbf{z}_i\mathbf{z}_j + e_{ij} > 0; \\ Y_{ij} &= 0, \text{ otherwise}, \end{aligned} \tag{2.6}$$

where $(\beta_{kl,0},\ \boldsymbol{\beta_{1,kl}}, \boldsymbol{\beta_{2,kl}})$ are the parameters governing the edge formation probability, and $\mathbf{z}_i\mathbf{z}_j$ is a shorthand (with some abuse of notation) for interaction terms of $\mathbf{z}_i$ and $\mathbf{z}_j$. Because we deal with an undirected graph, we further assume that the process (2.6) satisfies parameter symmetry over (k,l), i.e.,

$$\begin{aligned} \beta_{0,kl} &= \beta_{0,lk}, \\ \boldsymbol{\beta_{1,kl}} &= \boldsymbol{\beta_{1,lk}}, \\ \boldsymbol{\beta_{2,kl}} &= \boldsymbol{\beta_{2,lk}}, \end{aligned} \tag{2.7}$$

and that $e_{ij} \sim$ i.i.d. $\mathcal{N}(0,1)$.

To simplify notation in the likelihood function in Section 2.4, we will denote the conditional probability of two vertices i and j forming an edge as η_{ij} and express it in a compact form as:[55]

$$\begin{aligned} \eta_{ij}(X_i, X_j) &= \mathbb{P}(Y_{ij} > 0 | X_i, X_j, \mathbf{z}_i, \mathbf{z}_j) \\ &= \Phi\left(\boldsymbol{\gamma}(X_i, X_j, \mathbf{z}_i, \mathbf{z}_j)'\boldsymbol{\beta}\right), \end{aligned} \tag{2.8}$$

[54]For example, the naive Bayes model is widely used as a spam filter to classify emails into normal vs. spam emails, and it assumes that conditional on text class, the presence of words are independent.

[55]Note that we omit the covariates $\mathbf{z}_i$ and $\mathbf{z}_j$ in the arguments of expression (2.8) for η_{ij} for notational convenience.

where

$$\begin{aligned}
\boldsymbol{\gamma}(X_i, X_j, \mathbf{z}_i, \mathbf{z}_j) &= \mathbf{D}(X_i, X_j) \bigotimes [1,\ \mathbf{z}_i{+}\mathbf{z}_j,\ \mathbf{z}_i\mathbf{z}_j]',\\
\mathbf{D}(X_i, X_j) &= \left[\mathbb{1}_{(X_i=k, X_j=l)\vee(X_i=l, X_j=k)}\right]_{1\le k\le l\le m},\\
\boldsymbol{\beta} &= [\beta_{0,kl},\ \boldsymbol{\beta_{1,kl}},\ \boldsymbol{\beta_{2,kl}}]'_{1\le k\le l\le m}.
\end{aligned}$$

Weights of the Edge. Once two vertices form an edge, we assume that the weight of their edge, which in our application will correspond to the total amount of financial transactions between the two vertices, is drawn from a Q-valued discrete distribution with probabilities that depend on ideological proximity. Specifically, conditional on $Y_{ij} > 0$, if $X_i = k, X_j = l$, the Y_{ij} takes values $(w_1, ..., w_Q)$ with probabilities

$$\mathbf{h}_{kl} \equiv (h_{kl,1}, ..., h_{kl,Q}) \in \Delta^{Q-1}. \tag{2.9}$$

In our estimation we will pre-fix Q and the values $(w_1, ..., w_Q)$. Again because we deal with an undirected graph, we impose the natural symmetry assumption that

$$\mathbf{h}_{kl} = \mathbf{h}_{lk}. \tag{2.10}$$

Discussions

First, we would like to note that the edge formation process as specified by (2.6) can be interpreted as resulting from a model of matching. Two committees decide whether to establish a financial connection based on their joint surplus from forming a match. The surplus has a deterministic component $\beta_{0,kl} + \boldsymbol{\beta_{1,kl}}(\mathbf{z}_i{+}\mathbf{z}_j) + \boldsymbol{\beta_{2,kl}}\mathbf{z}_i\mathbf{z}_j$ as well as a stochastic component e_{ij}. Parameters differ by ideology pair (k, l). $\beta_{kl,0}$ captures the direct effect of ideology proximity, and $\boldsymbol{\beta_{1,kl}}$ and $\boldsymbol{\beta_{2,kl}}$ capture the effect of committee specific characteristics interacting with ideology proximity. In other words, ideology influences the edge formation probability through both the constant term and the coefficients.[56]

Second, financial and institutional characteristics $\mathbf{z}_i$ and $\mathbf{z}_j$ are included in the specification (2.6) not only because they may be important factors in the PCs' contribution decisions, but also, technically, it is our strategy for *degree correction*. Without these covariates, the model is essentially the same as Snijders and Nowicki (1997), where edge formation probability is governed only by ideology proximity. In their model, variations in degree are attributed only to random shock and ideology proximity. Moreover, the implied degree distribution is a mixture of $m(m+1)/2$ binomial distributions. However, when the number of ideology categories is small, the model-implied degree distribution may not be able to capture the empirical degree distribution. Therefore, we include vertex specific characteristics to introduce richer heterogeneity in the edge formation probability. This is a variant of the degree correction strategy introduced in Peng and Carvalho (2015a) and Peng and Carvalho (2015b). In addition to proximity of the latent ideology, Peng and Carvalho (2015a) assume that the edge formation probability also depends on additional latent variables $\xi_i, \xi_j \in \mathbb{R}$, capturing vertex specific popularity. Peng and Carvalho (2015b) replace the latent popularity variables with the quantile ranks of the vertices' observed degrees $q_i, q_j \in \{1, 2, ..., Q\}$. Our paper differs from Peng and Carvalho (2015b) in that the characteristics we use do not involve any features of the network, rendering a much cleaner model with which to perform statistical inference because the potential problem of endogeneity is avoided. Compared to Peng and Carvalho (2015a), our paper has a faster convergence rate and lower computational intensity because it does not require

[56] In an empirical application, additional constraints can be added: for example, restricting some coefficients in $\boldsymbol{\beta}_{\mathbf{1},\boldsymbol{kl}}$ and $\boldsymbol{\beta}_{\mathbf{2},\boldsymbol{kl}}$ to be 0, or to be the same for different ideology pairs. We provide the details of the additional restrictions in Section 2.6.

inferring about additional latent variable ξ_i.

Finally, note that conditional on the latent ideologies of the vertices, our model of network formation is equivalent to a pair-wise matching model. However, since the vertices' latent ideologies are not known, our edge formation process as specified in (2.6) allows for correlation *conditional on observables* $\left\{\hat{X}_i\right\}_{i\in\mathcal{V}^o}$ and $\mathbf{z}$. Moreover, the correlation is spread all over the connected component of the network via the latent ideologies. As a result, our network is not decomposable based on observables.

2.3 Data and Descriptive Statistics

Three data sets are used in our study, two of which are administrative data sets from the Federal Election Commission (FEC) in 2003-2004 election cycle: the " Committee Master File", and the "Contributions to Committees from Committees" data; and the third data set is collected from `OpenSecrets.org`.[57]

The "Committee Master File" contains basic information about all the PCs registered with the FEC. The PCs in this data set include federal political action committees (PACs), party committees, campaign committees for presidential, house and senate candidates, as well as groups or organizations such as lobbyists or fund raisers. This data set also contains information on the PC's geographical location, institutional characteristics, and the self-reported party affiliation, if available.

The "Contributions to Committees from Committees" data contains the *universe* of the contribution records between PCs. We observe the universe of records because the campaign finance laws mandate the disclosure of all transactions and expenditure related to federal election activities. In this data set, for each contribution, it lists the

[57]The website is maintained by the Center for Responsive Politics.

contributor, recipient, and the amount of contribution. We will use this data set to construct the *complete* network of financial transactions among the PCs.

Finally, the `OpenSecrets.org` data set contains the industry categorization for the PCs. The data sources corresponding to each of the variables we use are summarized in Table 2.1.

Variable	Data Source
$\mathbf{y}$	Contributions to Committees from Committees
$\hat{\mathbf{x}}$	Committee Master File
$\mathbf{z}$	Committee Master File (location and institutional characteristics), Contributions to Committees from Committees (imputed financial budget), `OpenSecrets.org` (industry categorization).

Table 2.1: Data Source

Vertex Characteristics

In the 2003-2004 election cycle, 5,858 PCs participated in contribution activities with other PCs and formed a financial network, among a total of 7,559 active PCs (defined as PCs with either positive total receipts or positive total disbursement).[58] We do not include in our study the PCs that do not participate in financial contributions with other PCs, though they could alternatively be viewed as isolated vertices in the network. This is not particularly worrisome because the PCs excluded from our study are financially insignificant. Among all PCs with positive receipts, the PCs in the financial network accounted for 96.83% of total amount of receipts; and among all

[58] Information on total receipts and disbursement is obtained from data sets "Candidate Summary (All Candidates)" and "PAC & Party Summary" released by the FEC. This information is missing for 89 PCs in the observed financial network. We define them as active, and use their budget in the financial network as a proxy for their total receipts and total disbursement.

PCs with positive disbursement, these PCs accounted for 96.87% of total amount of receipts.

Since the campaign finance laws do not mandate that political committees report their party affiliations, the Committee Master File has *incomplete* information on the PCs' political affiliation. In Table 2.2, we summarize the distribution of *reported* party affiliations of the PCs in the Committee Master File. Respectively 17.4% and 17.12% of the PCs reported to be Democratic and Republican, while 1.01% of the PCs reported to be Independent, and 0.85% of the PCs reported other affiliations such as Labor Party or Conservative Party. Importantly, 63.62% of the PCs did not report their party affiliations. In terms of financial significance, contributions sent by the PCs without self reported political affiliations accounted for 15.25% of the total amount of contributions among the PCs, and contributions received by them accounted for 37.38% of the total contributions.

Reported Affiliation	Frequency	% of PC	% of Contribution From	% of Contribution To
Democratic	1,019	17.40	45.25	34.74
Republican	1,003	17.12	38.64	25.78
Independent	59	1.01	0.53	1.63
Other	50	0.85	0.33	0.46
Missing	3,727	63.62	15.25	37.38
Total	5,858	100	100	100

Table 2.2: Tabulation of Reported Party Affiliation

In Table 2.3 we describe the non-ideological characteristics included in our analysis. Information on state and industry enables us to capture the effect of geographical and industrial proximity on political contribution. Figure 2.1 shows that the District of Columbia has the highest number of political committees (896), followed by California (575), Virginia (436), Texas (301), New York (284), Pennsylvania (273) and Illinois

Characteristics	Type	Range
State	Categorical	55 distinct categories
Industry	Categorical	46 distinct categories
Committee Type	Categorical	6 distinct categories
Committee Designation	Categorical	3 distinct categories
National	Dummy	{0,1}
Budget (in $1,000)	Continuous	[0, 104064.2]

Table 2.3: Non-ideological Characteristics
Note: States include unincorporated territories.

(226). In Table 2.4, we provide a coarse industrial breakdowns of the political committees. Besides "Other" , which aggregates many small industrial categories, Finance/Insurance/Real Estate is the most represented industry (9.12%), followed by Miscellaneous Business (6.96%), Health (5.43%) and Labor (5.19%). In our empirical analysis, we drop the ideological and party affiliated sectors, and use a finer breakdown for sectors such as Misc Business and Other. We provide a detailed description of the construction of the industry variable in Appendix 4.1.

Committee type, committee designation, and national dummy are institutional characteristics that are potentially related to contribution patterns. There are six distinct categories of committee type in the Committee Master File: House campaign, Senate campaign, Presidential campaign, qualified PAC, qualified Party and others.[59] Table 2.5 shows that 27.72% of the committees are campaign committees of either House, Senate or Presidential elections, 45.43% are qualified PACs, and 4.52% are qualified Party committees, and the rest ("Other") are mainly non-qualified committees.

[59] Qualified PACs need to have been in existence for six months and received contributions from 50 people and made contributions to five federal candidates. Qualified party committees need to have been in existence for at least six months and received contributions from 50 people or are affiliated with another party committee that meets these requirements.

Sector	All PCs		Report Dem		Report Rep		Report Ind		Missing/Report Other	
	Number	%	Number	%	Number	%	Number	%	Number	%
Agribusiness	285	4.87	0	0	0	0	4	6.78	281	7.44
Communic/Electronics	186	3.18	0	0	0	0	3	5.08	183	4.85
Construction	124	2.12	0	0	0	0	1	1.69	123	3.26
Defense	54	0.92	0	0	0	0	1	1.69	53	1.40
Energy/Natural Resource	274	4.68	1	0.10	0	0	3	5.08	270	7.15
Finance/Insurance/ Real Estate	534	9.12	0	0	0	0	6	10.17	528	13.98
Health	318	5.43	0	0	0	0	4	6.78	314	8.31
Labor	304	5.19	1	0.10	0	0	9	15.25	294	7.78
Lawyers & Lobbyists	186	3.18	0	0	0	0	0	0	186	4.92
Transportation	157	2.68	0	0	0	0	2	3.40	155	4.10
Misc Business	408	6.96	0	0	0	0	6	10.17	402	10.64
Ideological/Single-Issue	2,404	41.04	782	76.74	812	80.96	19	32.20	791	20.94
Joint Candidate Committee	164	2.80	81	7.95	65	6.48	0	0	18	0.48
Party Committee	359	6.13	153	15.01	124	12.36	0	0	82	2.17
Other	24	0.41	0	0	0	0	1	1.69	23	0.61
Unknown	77	1.31	1	0.01	2	0.20	0	0	74	1.96
Total	5,858	100	1,019	100	1,003	100	59	100	3,777	100

Table 2.4: Industrial Sectors of Political Committees

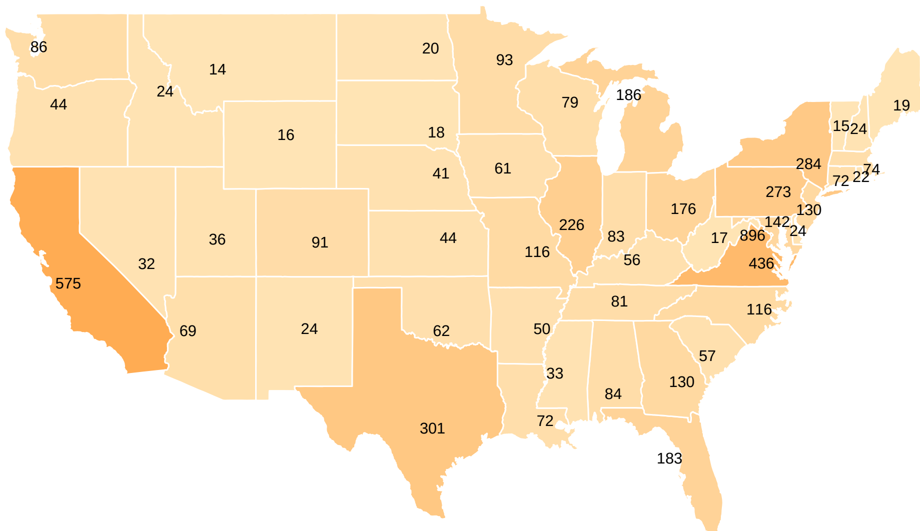

Figure 2.1: Geographical Distribution of Political Committees

In Table 2.6, we provide information about the committee designations. There are three distinct categories of committee designation in the Committee Master File: authorized by a candidate, joint fund-raiser, and others.[60] Table 2.6 shows that 1.89% of the committees are authorized by a candidate, 2.83% are joint fund-raisers, and the majority of the rest are either principal campaigns or committees not authorized by a candidate. We do not list principal campaigns separately because it would be redundant given the more detailed categorization in committee type. Finally, the national dummy listed in Table 2.3 takes value 1 if and only if the committee is one of the following six committees: the Democratic National Committee (DNC), the Democratic Senatorial Campaign Committee (DSCC), the Democratic Congressional

[60] A committee is designated as "*authorized by a candidate*" if it is authorized by a candidate in writing to receive contributions or make expenditures on behalf of the candidate, but is not her principal campaign committee. A committee is designated as "*a joint fundraiser*" if it is created by two or more candidates, PACs or party committees to share the costs of fundraising, and split the proceeds.

Campaign Committee (DCCC), the Republican National Committee (RNC), National Republican Senatorial Committee (NRSC), and National Republican Congressional Committee (NRCC).

We compute a PC's budget from its contribution record. It is constructed as the *sum of its contributions to* all other PCs within this election cycle. Typically, this amount is lower than a committee's total receipts or total disbursements. We use this measure because it is most powerful in explaining a PC's probability of making political contribution, and thus its total number of connections. In fact, the correlation between a PC's total receipts or disbursements and its number of contributions to other PCs is low. The distribution of budget is given in Table 2.7 and Figure 2.2. It has a wide range and a fat tail: 25.7% of the PCs have budget less than or equal to \$1,000, while 8% of the PCs have budget more than \$500,000.

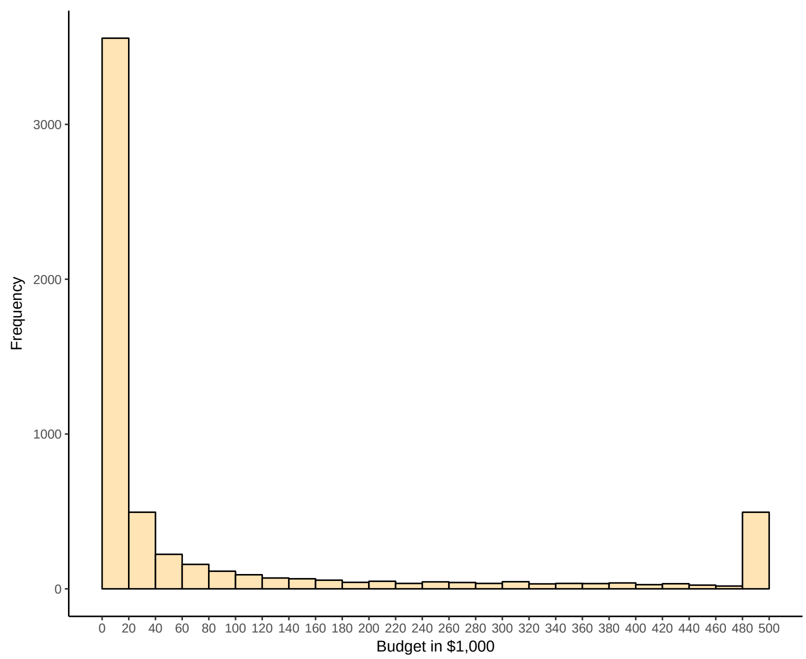

Figure 2.2: Distribution of Budget
Note: Observations with budget higher than \$500,000 are plotted at \$500,000.

	All PCs		Report Dem		Report Rep		Report Ind		Missing/Report Other	
Committee Type	Number	%	Number	%	Number	%	Number	%	Number	%
House	1,275	21.77	613	60.16	644	64.21	8	13.56	10	0.26
Senate	309	5.27	140	13.74	158	15.75	2	3.39	9	0.24
Presidential	40	0.68	20	1.96	9	0.90	6	10.17	5	0.13
Qualified PAC	2,661	45.43	20	1.96	13	1.30	43	72.88	2,585	68.44
Qualified Party	265	4.52	127	12.46	114	11.37	0	0	24	0.64
Other	1,308	22.33	99	9.72	65	6.48	0	0	1,144	30.29
Total	5,858	100	1,019	100	1,003	100	59	100	3,777	100

Table 2.5: Types of the Political Committees

Note: The 22.33% other committees mainly consist of non-qualified PACs and non-qualified Party committees.

	All PCs		Reported Dem		Reported Rep		Reported Ind		Missing/Reported Other	
Committee Type	Freq	%	Freq	%	Freq	%	Freq	%	Freq	%
Authorized by a Candidate	111	1.89	42	4.12	65	6.48	4	6.78	0	0
Joint Fundraiser	166	2.83	81	7.95	66	6.58	0	0	19	0.50
Other	5,581	95.27	896	87.93	872	86.94	55	93.22	3,758	99.50
Total	5,858	100	1,019	100	1,003	100	59	100	3,777	100

Table 2.6: Committee Designation

Note: The 95.27% other committees mainly consist of principal campaigns and committees that are not authorized by a candidate.

Quantile	All PCs	Report Dem	Report Rep	Report Ind	Missing/Report Other
Min	0.00	0.00	0.00	0.00	0.00
25%	1.00	2.98	4.00	5.00	0.70
50%	10.00	32.75	48.07	20.00	6.08
75%	75.90	353.30	456.15	88.87	30.00
Max	104,064.19	104,064.19	55,873.70	2,386.53	9,180.75
Obs.	5,858	1,019	1,003	59	3,777

Table 2.7: Quantiles of Budget in $1,000

Descriptive Statistics of the Financial Transactions Network

The Contributions from Committees to Committees data set records 411,106 transactions among 5,858 political committees in the 2003-2004 election cycle. Figure 2.3 is a graphical representation of the network using graphing software Gephi.[61] Each vertex represents a PC, and the color of the vertex represents reported affiliation: *blue* for Democratic, *red* for Republican, *green* for Independent, and *yellow* for missing/other. For the purpose of this graph, we collapse multiple contributions with the same direction between any pair of PCs into one edge: each edge represents a *directed* financial connection; and we use the contributor's color to represent the color of the edge. Thus there are a total of 164,529 edges in Figure 2.3. This network has 20 disconnected components.[62] Nevertheless, the component in the center is disproportionately large with 5,806 vertices. The subsequent description focuses on this giant component.

Network Level Statistics. In the rest of the analysis we will consider the network as undirected, thus we further collapse transactions with different directions between any pair of PCs. We derive a financial network with 5,806 vertices and 145,406 edges.

[61] See https://gephi.org.

[62] A component is a maximum connected sub-network.

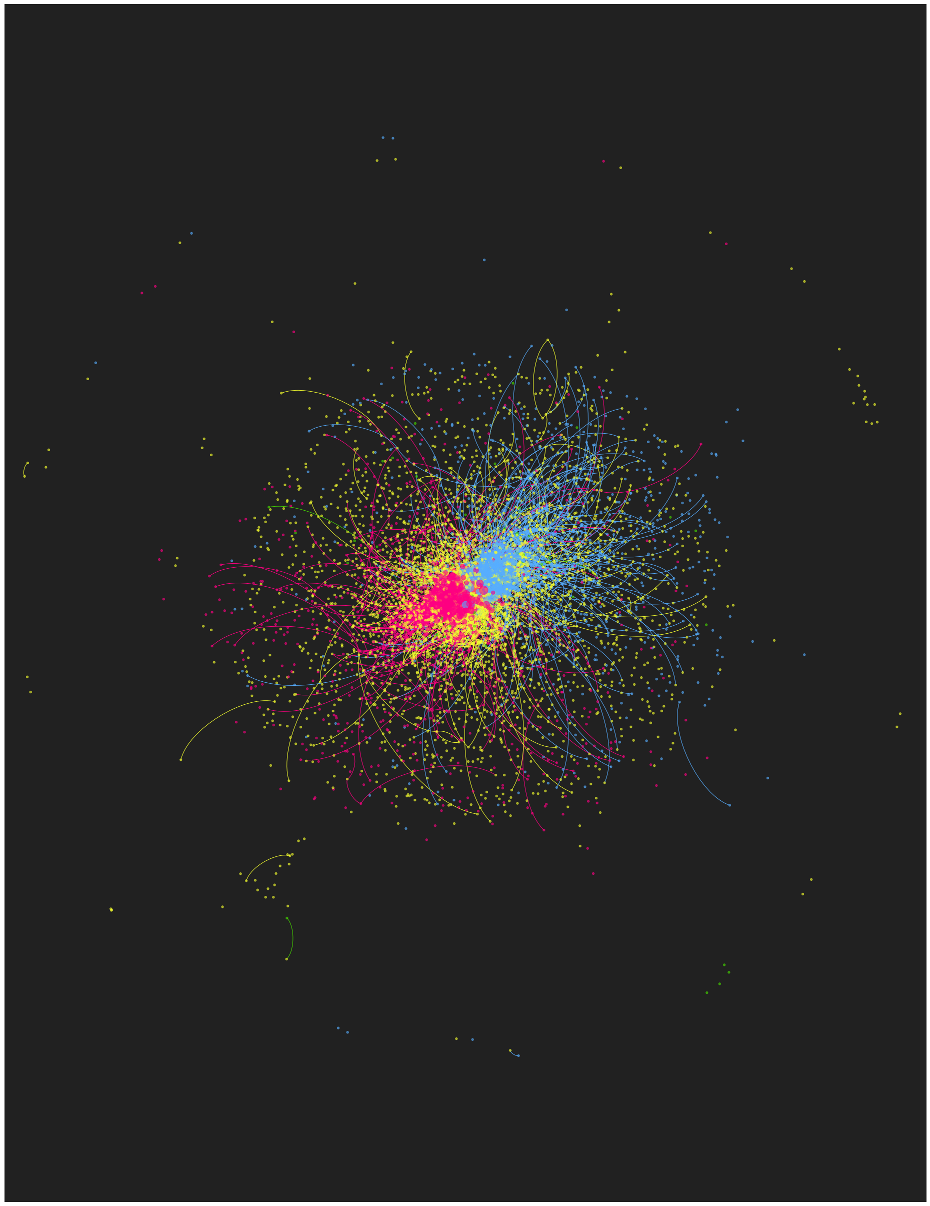

Figure 2.3: Political Contribution Network

Note: Vertex color represents reported affiliation: blue for Democratic, red for Republican, green for Independent, and yellow for missing. Edge color is the same as contributor's color.

Table 2.8 presents some of the key statistics for the financial transaction network among the PCs. First, the network is sparse, yet well-connected. On the one hand, the average degree is 50.09, i.e. on average, a PC is connected to only 0.86% of other PCs.[63] This indicates sparsity because the number of edges is only a small fraction of all the vertex pairs. On the other hand, it has a diameter of 10, and an average distance of 2.99.[64] Both statistics indicate that the network is well-connected, which poses a challenge for our study. Given the connectedness, there is no straightforward approach to structurally decompose this giant component into separate sub-networks, despite the visual patterns of clustering. Moreover, there is no other natural way of decomposition. A common practice in the network applications is to partition the full network into geographically disjoint sub-networks and assume no interaction across sub-networks. However, this practice is not suitable in our application because political contributions are not concentrated at local levels.

Number of vertices	5,806
Number of edges	145,406
Average degree	50.09
Diameter	10
Average distance	2.99

Table 2.8: Network Statistics

Distributions of Degrees and Edge Weights. To explore beyond the network-level summary statistics, we further investigate the distribution of degrees. Figure 2.4 shows that the degree distribution has a large spread and a fat tail: the highest

[63] *Degree* is a vertex-level statistics. It is the number of direct connections a vertex has. Following the notation introduced in the previous section, vertex i's degree is given by $d_i = \sum_{j \neq i} \mathbb{1}(y_{ij} > 0)$.

[64] The *diameter* of a network is calculated as the maximum length of the (finite) shortest paths among vertex pairs. The *average distance* of a network is calculated as the average length of the (finite) shortest paths among vertex pairs.

degree is as large as 978. An important feature is that the shape of this distribution is inconsistent with a mixture of a small number of binomial distributions–an implication in the standard stochastic block model. To capture the observed degree distribution, we introduce rich heterogeneity in our model. Specifically, we include budget whose distribution has similar patterns to account for the variations in degree.

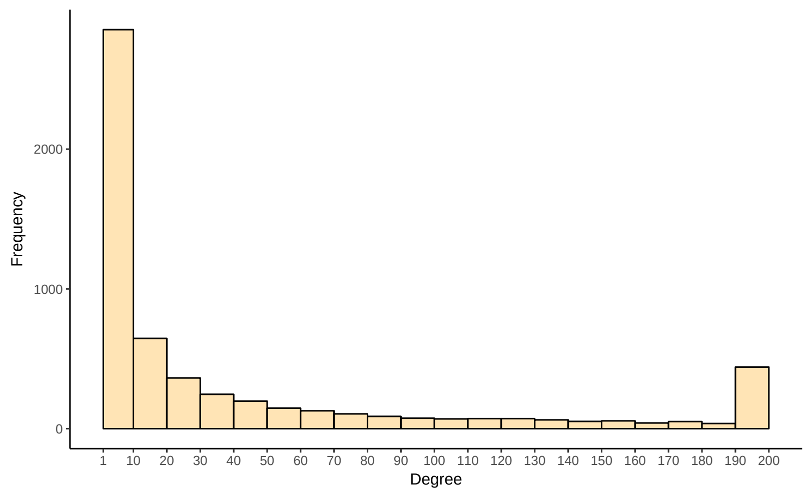

Figure 2.4: Histogram of degree
Note: Observations with degree higher than 200 are plotted at 200.

Additionally, we study the sub-networks induced by reported Democratic, Republican and Independent PCs, and analyze connection patterns conditional on reported affiliations. In Figure 2.5, we have three graphs for self-reported Democratic, Republican, and Independent PCs respectively. In each graph, we present the distribution of the PCs' numbers of connections with the three groups of self identified PCs. Self-reported Democratic and Republican PCs show evidence of homophily. On the one hand, many PCs are financially connected with PCs affiliated to same party, and a sizable of them have more than 20 such connections; on the other hand, only a small number of PCs are financially connected with PCs affiliated to the other party, and most of them have less than 10 such connections. Self-reported Independent PCs show similar connection pattern with the self-reported Democratic vs. Republican PCs. Many are financially connected with both parties, and a sizable of them have more than 20 such connections. A small fraction are financially connected with other self-reported Independent PCs, and most of them have less than 10 such connections - partially due to the small number of self-reported Independent PCs. Note that we cannot claim, based on the connection patterns between PCs with self reported affiliations, that this is an evidence of heterophily. It is possible that they are connected with a large number of Independent PCs without self-reported affiliations. Furthermore, even if Independent PCs exhibit heterophily, this behavior does not invalidate either our model or estimation. Identification only requires that PCs with different political ideologies have different contribution patterns. This is at least true for the PCs with self-reported affiliation, which is reassuring.

Finally, in order to infer about the unknown ideologies from the PCs with self reported ideologies, an implicit assumption in our model is that the PCs without self reported affiliations do not act systematically differently from the PCs with self reported affiliations. Therefore, we reproduce the degree distribution in Figure 2.6 with the composition of each bar marked by color. The degree distribution of PCs without self reported affiliations is similar to that of the PCs with self reported affiliations. This evidence is consistent with the assumption, though insufficient. In Section 2.6, we will present more evidence on this after we have presented our estimation results.

The financial network also provides information on transfer amount (the sum of the contribution amount) between pairs of connected PCs. Its distribution is given in Table 2.9 and Figure 2.7. Most of them are small because of the regulations on

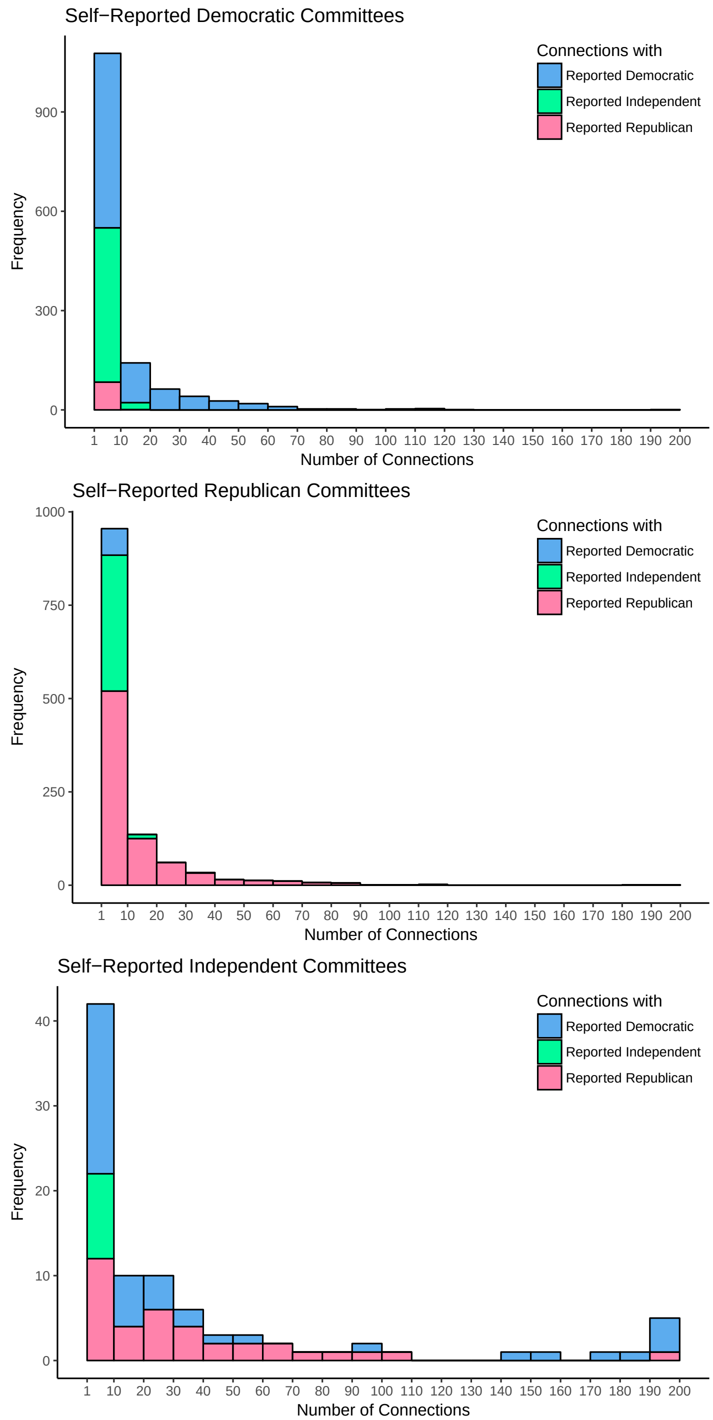

Figure 2.5: Distributions of Number of Connections with Different Committees

Note: These figures only include observations with positive number of connections. Observations with more than 200 connections are plotted at 200.

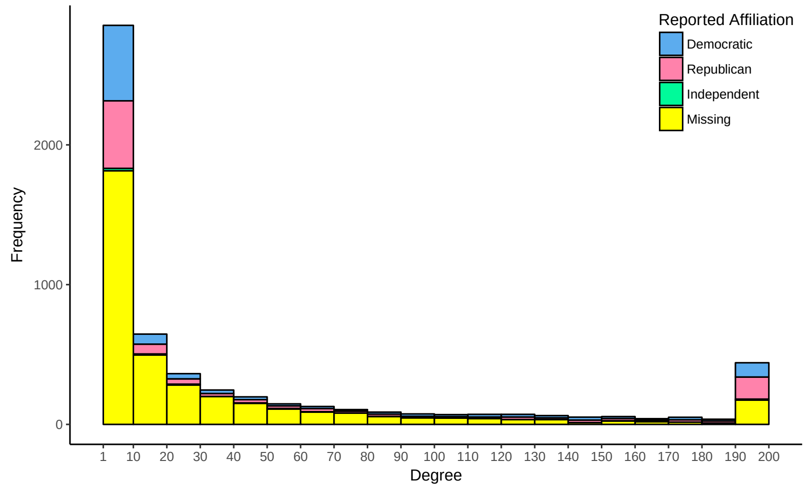

Figure 2.6: Decomposition of Degree Distribution
Note: The same as Figure 2.4, but the composition of each bar, i.e., the composition of PCs with a certain rage of degrees, is presented with different colors.

contribution limits (detailed description in Appendix 4.2). In addition, we present the distribution of transfer amount conditional on reported ideology in Table 2.9 and Figure 2.8. On average, the transfer amount is high between PCs with the same reported affiliation, and low for the other cases. A caveat in interpreting these statistics is that we exclude all the contributions involving PCs without self reported affiliations, so they do not necessarily give the full picture of the contribution pattern.

Issues with Naive Alternative Methods

In this subsection, we briefly discuss the issues with some naive alternative methods that we have attempted, and explain why they are invalid. We first define

Quantile	All	Dem,Dem	Rep,Rep	Ind, Ind	Other/Missing,Other/Missing	
Min	1.00	6.00	10.00	5,000.00	11.00	
25%	1,000.00	1,000.00	1,000.00	10,000.00	2,000.00	
50%	2,000.00	1,000.00	1,000.00	10,000.00	5,000.00	
75%	5,000.00	2,000.00	4,028.00	29,143.00	10,000.00	
Max	47,190,000.00	47,190,000.00	22,542,215.00	223,000.00	12,367,982.00	
Mean	8,990.74	76,034.83	58,189.32	48,920.67	9,544.85	
Obs.	145,406	5,582	5,151	6	13,979	
Quantile	Dem,Rep	Dem,Ind	Dem,Other/Missing	Rep,Ind	Rep,Other/Missing	Ind,Other/Missing
Min	10.00	51.00	1.00	50.00	2.00	5.00
25%	1,000.00	1,500.00	1,000.00	1,000.00	1,000.00	2,000.00
50%	1,000.00	4,000.00	2,000.00	2,000.00	2,000.00	5,000.00
75%	2,000.00	7,500.00	5,000.00	5,000.00	4,000.00	10,000.00
Max	150,000.000	75,000.00	2,966,933.00	60,000.00	1,638,000.00	1,000,000.00
Mean	9,059.11	5,209.31	3,833.99	3,677.08	3,468.98	18,441.87
Obs.	121	2,238	48,132	1,359	68,281	557

Table 2.9: Quantiles of Transfer Amount (in \$1.00)

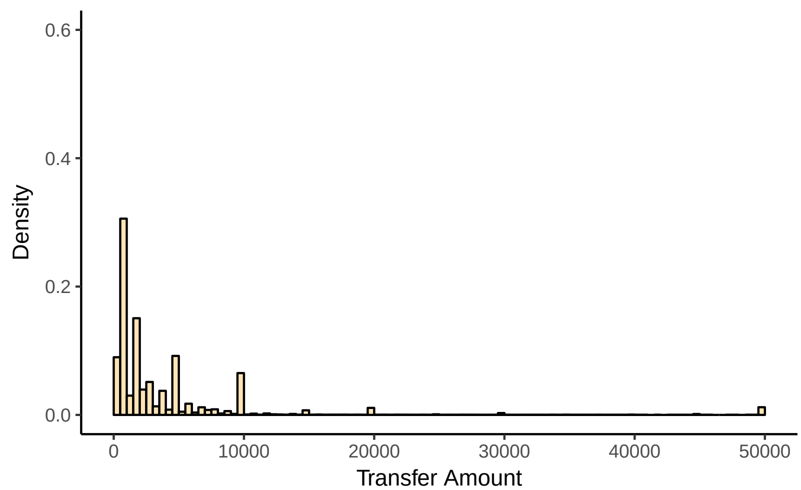

Figure 2.7: Empirical Distribution of Transfer Amount

Note: Bin size is 500. Observations with transfer amount higher than $50,000 are plotted at $50,000.

$$\hat{x}_i = \begin{cases} -1 \text{ if committee } i \text{ is reported to be Democratic} \\ \quad 0 \text{ if committee } i \text{ is reported to be Independent} \\ \quad 1 \text{ if committee } i \text{ is reported to be Republican.} \end{cases}$$

The naive method tries to find $\mathbf{x}^*$ solving the following fixed point problem as a solution to the ideology recovery problem:

$$\begin{bmatrix} \hat{\mathbf{x}} \\ \mathbf{x}^* \end{bmatrix} = \text{sign}\left(\mathbf{y}\begin{bmatrix} \hat{\mathbf{x}} \\ \mathbf{x}^* \end{bmatrix}\right). \tag{2.11}$$

In other words, a PC is Democratic (Republican) if it is connected with more Democratic (Republican) PCs than Republican (Democratic) ones, or Independent if it is connected with an equal number of Democratic and Republican PCs. However, neither existence nor uniqueness of the solution is guaranteed. We attempted to solve this problem by iteration method, but failed. The following example in Figure 2.9 illustrates the reason. According to the categorization rule described above, the

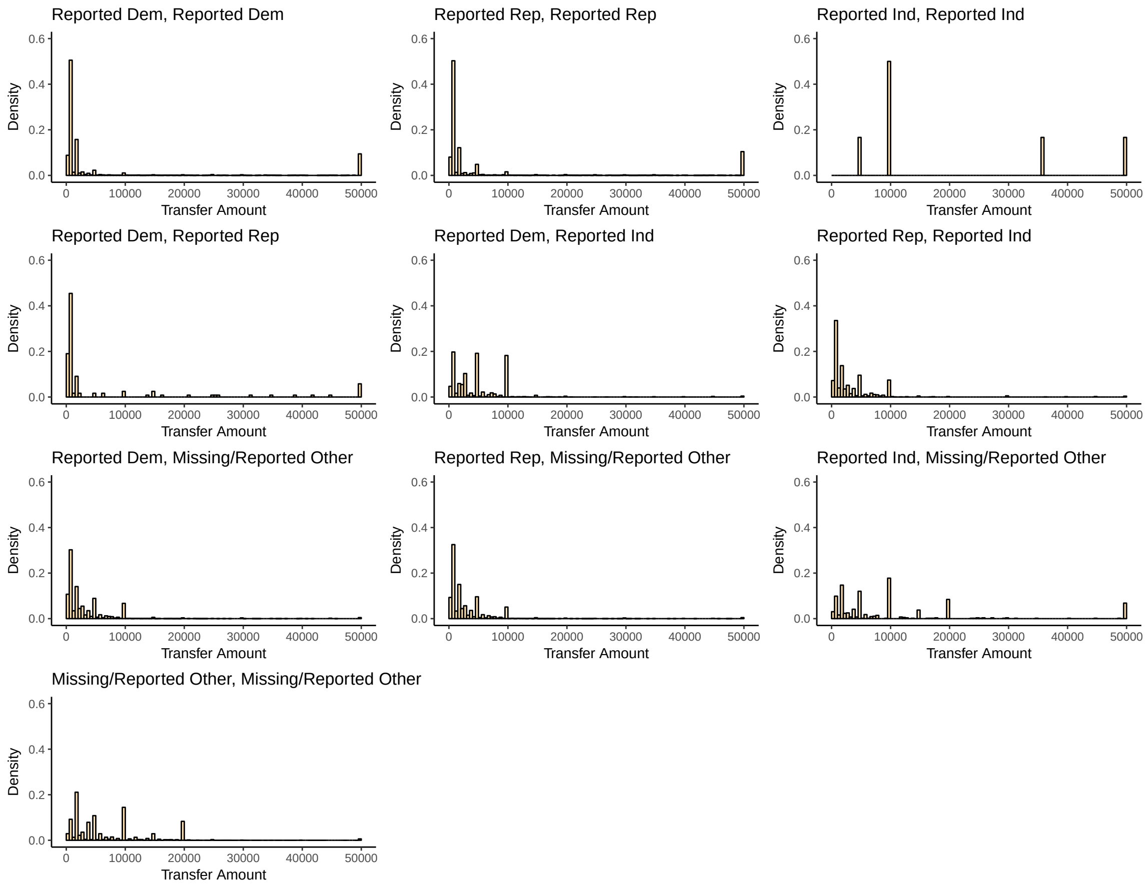

Figure 2.8: Empirical Distribution of Transfer Amount Conditional on Self-Reported Ideology

Note: Bin size is 500. Observations with transfer amount higher than \$50,000 are plotted at \$50,000.

two "unknown" vertices should be assigned Democratic. However, this assignment generates inconsistency in the "Republican" vertex's behavior: as a Republican PC, it is connected with two Democratic PCs. There does not exist a categorization which can reconcile such inconsistency, so the naive method proposed in (2.11) does not guarantee a well-defined solution.

Since the solution concept above requires too strong a coherency in categorization,

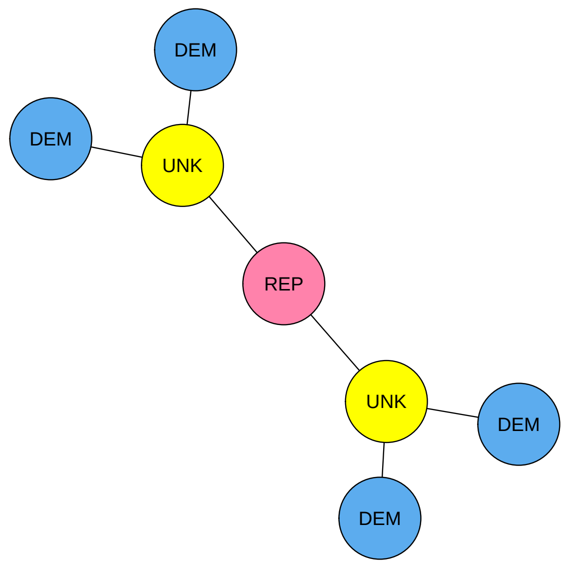

Figure 2.9: Non-Existence of Solution with Naive Method

a less restrictive method was also attempted:

$$x_i = \sum_{j \in \mathcal{V}^o} y_{ij} \hat{x}_j. \tag{2.12}$$

In this case, a PC's ideology is defined by its connected PCs with self reported affiliation. There are two major problems with this method. First, when a PC is only connected to PCs with unreported affiliations, its ideology is not defined. If we try to address this problem by iteratively applying the equation above, we go back to the previous method. Second, categorization has very low precision for PCs connected with few PCs with self reported affiliations and mostly with PCs with unknown affiliations. The poor performance of naive methods necessitate the use of a more sophisticated method.

2.4 Estimation

Given the model description, the likelihood of $\mathbf{y}, \hat{\mathbf{x}}$ conditional on $\mathbf{x}, \mathbf{z}$ is given by

$$\begin{aligned} & \mathbb{P}\left(\mathbf{y}, \hat{\mathbf{x}} | \mathbf{x}, \mathbf{z}; \epsilon, \boldsymbol{\beta}, \mathbf{h}\right) \\ = & \mathbb{P}\left(\hat{\mathbf{x}} | \mathbf{x}; \epsilon\right) \mathbb{P}\left(\mathbf{y} | \mathbf{x}, \mathbf{z}; \boldsymbol{\beta}, \mathbf{h}\right) \\ = & (1-\epsilon)^{n_t} \left(\tfrac{\epsilon}{m-1}\right)^{n_e} \textstyle\prod_{1 \leq i < j \leq n} \left[\eta_{ij}(x_i, x_j) h_{x_i x_j}(y_{ij})\right]^{\mathbb{1}(y_{ij}>0)} \left[1 - \eta_{ij}(x_i, x_j)\right]^{\mathbb{1}(y_{ij}=0)}, \end{aligned} \tag{2.13}$$

where $n_t = \sum_{i \in \mathcal{V}^o} \mathbb{1}(x_i = \hat{x}_i)$ is the number of vertices in $\mathcal{V}^o$ whose $\hat{x}_i$'s coincide with x_i's, $n_e = \sum_{i \in \mathcal{V}^o} \mathbb{1}(x_i \neq \hat{x}_i)$ is the number of vertices in $\mathcal{V}^o$ whose $\hat{x}_i$'s differ from x_i's, and $h_{x_i x_j}(y_{ij}) = h_{x_i x_j, q}$ if $y_{ij} = w_q$ where $h_{x_i x_j, q}$ is defined in (2.9).

We use the Maximum A Posteriori (MAP) estimator to recover the latent ideology. It is a Bayesian estimator that equals the mode of the posterior probability. Specifically, it solves

$$\max_{\mathbf{x} \in \{1, \dots, m\}^n} \mathbb{P}\left(\mathbf{y}, \hat{\mathbf{x}} | \mathbf{x}, \mathbf{z}; \epsilon, \boldsymbol{\beta}, \mathbf{h}\right) \mathbb{P}\left(\mathbf{x}; \boldsymbol{\theta}\right). \tag{2.14}$$

Note that the Maximum Likelihood Estimator (MLE) solves

$$\max_{\mathbf{x} \in \{1, \dots, m\}^n} \mathbb{P}\left(\mathbf{y}, \hat{\mathbf{x}} | \mathbf{x}, \mathbf{z}; \epsilon, \boldsymbol{\beta}, \mathbf{h}\right), \tag{2.15}$$

and that MAP is equivalent to MLE under uniform prior $\boldsymbol{\theta} = (\frac{1}{m}, \dots, \frac{1}{m})$.

We now argue for the validity of the MAP estimator, and propose a Bayesian algorithm to obtain an approximate solution. Theoretically, statistical inference is non-standard in our model. First of all, we have only one observation of the network, i.e. one realization of $\mathbf{y}, \hat{\mathbf{x}}$. Second, the number of parameters (the number of latent political ideologies) grows with the network size. Therefore, canonical asymptotic theory is not applicable. For example, law of large numbers cannot be directly applied. As a result, in the recent literature, new concepts and tools are introduced to study this problem. The following subsection summarizes the key concepts and results mostly related to our study.

Threshold for Exact Recovery

In this subsection we describe the theoretical results in Yun and Proutiere (2016) that justify our estimation using one observation of the network. The standard stochastic block model was first introduced in Snijders and Nowicki (1997) and Nowicki and Snijders (2001), and its exact recovery problems were studied in Mossel et al. (2014), Abbe et al. (2014), and Abbe and Sandon (2015). The labeled stochastic block model introduced weights on edges, and its exact recovery problems were studied in Jog and Loh (2015) and Yun and Proutiere (2016). First, we provide a formal definition of the labeled stochastic block model.

Definition 1 (Labeled Stochastic Block Model LSBM $(n, \boldsymbol{\theta}, \frac{\log(n)}{n}\boldsymbol{W})$). The Labeled Stochastic Block Model generates an n-vertex random graph with community affiliation $\boldsymbol{X}$ and weighted adjacency matrix $\boldsymbol{Y}$ according to the following process. Each vertex is assigned a community affiliation $X_i \in \{1, 2, ..., m\}$ independently under probability $\boldsymbol{\theta} = (\theta_1, ..., \theta_m) \in \Delta^{m-1}$. Conditional on community affiliations, the edges are drawn independently. The edges Y_{ij}'s take discrete values $\{0, w_1, w_2, ...w_Q\}$ where 0 represents no edge and w_q represents an edge with the specified (non-zero) weight. The distribution of edges is governed by an m-by-m-by-Q matrix $\boldsymbol{W}$. Specifically, vector $\boldsymbol{W}(k, l; \cdot)$ characterizes edge distribution between a pair of vertices in communities k and l:

$$\begin{aligned}\mathbb{P}(Y_{i,j} = w_q | X_i = k, X_j = l) &= \frac{\log(n)}{n} W(k,l;q) \quad \text{for } q = 1,..,Q;\\ \mathbb{P}(Y_{i,j} = 0 | X_i = k, X_j = l) &= 1 - \frac{\log(n)}{n} \sum_{q=1}^{Q} W(k,l;q).\end{aligned} \tag{2.16}$$

Note that $\boldsymbol{W}$ is not indexed by the network size n, i.e. it does not change with n. This implies that the distribution described above will change with n. More precisely, the probability of having an edge scales as $\Theta(\frac{\log(n)}{n})$ and the degree scales as $\Theta(\log(n))$. This logarithmic growth rate of degree with respect to the network size is called the logarithmic degree regime. The literature on exact recovery studies this regime because it is dense enough that the graph is connected with high probability; yet it is still sparse enough that the conditional independence condition yields asymptotic independence of the failures of the component-MAP for different vertices.

Next, we provide the definition of exact recovery. Exact recovery is an asymptotic requirement in the context of the SBM - a counterpart of consistency in the classical statistical problems.[65]

Definition 2 (Exact Recovery). Exact recovery is solved if there exists an algorithm such that $\mathbb{P}(\boldsymbol{X}^{est} = \boldsymbol{X}) \to 1$ as $n \to \infty$ where $\boldsymbol{X}^{est}$ is the estimated community affiliation.[66]

In other words, exact recovery requires that for a large enough network, the probability of correctly recovering the entire community structure (i.e. no misclassification) is almost 1. The most promising estimator to solve exact recovery is the MAP estimator

[65] Exact recovery is sometimes referred to as strong consistency, reflecting the resemblance to consistency.

[66] The equivalence is up to group permutation of $\boldsymbol{X}^{est}$ with respect to community names.

because it minimizes $\mathbb{P}(\boldsymbol{X}^{est} \neq \boldsymbol{X})$ - if MAP fails in solving exact recovery, no other algorithm can succeed (see, e.g., Abbe (2017)).

In order to describe the condition for exact recovery, we first define the Chernoff-Hellinger (CH) divergence (Abbe and Sandon (2015)).

Definition 3 (Chernoff-Hellinger divergence $D(\boldsymbol{\theta}, \boldsymbol{W})$).

$$D(\boldsymbol{\theta}, \boldsymbol{W}) = \min_{k,l:k\neq l} D_{L+}(\boldsymbol{\theta}, \boldsymbol{W}(k), \boldsymbol{W}(l)) \tag{2.17}$$

where $D_{L+}(\boldsymbol{\theta}, \boldsymbol{W}(k), \boldsymbol{W}(l))$ is given by

$$\begin{aligned} &D_{L+}(\boldsymbol{\theta}, \boldsymbol{W}(k), \boldsymbol{W}(l)) \\ &= \max_{\lambda\in[0,1]} \textstyle\sum_{q=1}^{Q} \sum_{j=1}^{m} \theta_j[(1-\lambda)W(k,j;q) \\ &+\lambda W(l,j;q) - W(k,j;q)^{1-\lambda} W(l,j;q)^{\lambda}] \end{aligned} \tag{2.18}$$

Intuitively, $D_{L+}(\boldsymbol{\theta}, \boldsymbol{W}(k), \boldsymbol{W}(l))$ measures the difference in connection patterns between a pair of communities, k and l; and thus the CH-divergence $D(\boldsymbol{\theta}, \boldsymbol{W})$ is the minimum of such difference between any pair of distinct communities. In the following, we explain more precisely the meaning of difference in connection pattern. Note that mathematically $\theta_j[(1-\lambda)W(k,j;q) + \lambda W(l,j;q) - W(k,j;q)^{1-\lambda}W(l,j;q)^{\lambda}]$ measures the difference between $\theta_j W(k,j;q)$ and $\theta_j W(l,j;q)$. Moreover, $\theta_j W(k,j;q)\log(n)$ gives, for a vertex in community k, the expected number of w_q-weighted edges with community j; and $\theta_j W(l,j;q)\log(n)$ gives, for a vertex in community l, a similar number. Therefore, $\theta_j[(1-\lambda)W(k,j;q) + \lambda W(l,j;q) - W(k,j;q)^{1-\lambda}W(l,j;q)^{\lambda}]$ measures the difference between communities k and l in terms of the number of w_q-weighted edges with community j. Finally, summing over different communities j and different edge weights w_q delivers the expression in (2.18). $D_{L+}(\boldsymbol{\theta}, \boldsymbol{W}(k), \boldsymbol{W}(l))$ is non-negative.

Larger value of $D_{L+}(\boldsymbol{\theta}, \boldsymbol{W}(k), \boldsymbol{W}(l))$ represents larger difference in connection patterns between communities k and l, and $D_{L+}(\boldsymbol{\theta}, \boldsymbol{W}(k), \boldsymbol{W}(l))$ is zero if and only if the two communities have identical connection patterns, i.e., $W(k, j; q) = W(l, j; q), \forall j, q$.

We further illustrate the definition of CH divergence in a special case of a homogeneous model (Jog and Loh (2015)), where its expression is significantly simplified. In a homogeneous model, vertices are assigned to different communities with equal probabilities; and the distribution of an edge only depends on whether the pair of vertices belong to the same communities:

$$\theta_j = \frac{1}{m}, \forall j; \tag{2.19}$$

$$W(k, l; \cdot) = \begin{cases} W_{within}(\cdot) \text{ if } k = l \\ W_{between}(\cdot) \text{ if } k \neq l. \end{cases} \tag{2.20}$$

Under homogeneity, the CH divergence reduces to the Hellinger divergence (corresponding to $\lambda = \frac{1}{2}$), measuring the difference in within-community and between-community connection patterns:

$$\begin{aligned}
&D(\boldsymbol{\theta}, \boldsymbol{W}) \\
=&D_{L+}(\boldsymbol{\theta}, \boldsymbol{W}(k), \boldsymbol{W}(l)) \quad \forall k \neq l \\
=&\max_{\lambda \in [0,1]} \sum_{q=1}^{Q} \sum_{j=1}^{m} \theta_j [(1-\lambda) W(k, j; q) + \lambda W(l, j; q) - W(k, j; q)^{1-\lambda} W(l, j; q)^{\lambda}] \\
=&\max_{\lambda \in [0,1]} \frac{1}{m} \sum_{q=1}^{Q} \sum_{j \in \{k, l\}} [(1-\lambda) W(k, j; q) + \lambda W(l, j; q) - W(k, j; q)^{1-\lambda} W(l, j; q)^{\lambda}] \\
=&\max_{\lambda \in [0,1]} \frac{1}{m} \sum_{q=1}^{Q} W_{within}(q) + W_{between}(q) - W_{within}(q)^{1-\lambda} W_{between}(q)^{\lambda} - W_{between}(q)^{1-\lambda} W_{within}(q)^{\lambda} \\
=&\frac{1}{m} \sum_{q=1}^{Q} (\sqrt{W_{within}(q)} - \sqrt{W_{between}(q)})^2.
\end{aligned} \tag{2.21}$$

The first equation follows symmetry, the second equation directly applies the definition in (2.18), the third equation uses (2.19) and $W(k, j; \cdot) = W(l, j; \cdot) = W_{between}(\cdot)$ $\forall j \neq k, l$, the fourth equation applies (2.20), and the last equation results from $\lambda = \frac{1}{2}$ being the maximizer.

Now we can state the main theoretical results. Combining Theorem 3 and Claim 4 in Yun and Proutiere (2016), we have:

Theorem 1 (Threshold for Exact Recovery). Exact recovery is solvable for LSBM $(n, \boldsymbol{\theta}, \frac{\log(n)}{n}\boldsymbol{W})$ if $D(\boldsymbol{\theta}, \boldsymbol{W}) > 1$.

The theorem shows that: if the difference in connection patterns of any pair of communities is large enough, we can correctly recover the entire community structure from one observed network with high probability. This theorem provides theoretical foundation for the use of the MAP estimator (because it is the "optimal" algorithm in terms of exact recovery), and it is very similar to the consistency results in classical statistics.

Some comment on the case where $D(\boldsymbol{\theta}, \boldsymbol{W}) < 1$ is useful. In this case, the MAP estimator fails exact recovery (i.e. has misclassification) with strictly positive probability. This result should not be interpreted as discouraging: although the probability of having misclassification does not vanish with the growth of the sample size, the misclassification rate defined as the proportion of vertices misclassified could still be low. In our Monte Carlo simulations, we observe that even when the CH divergence is below 1, we still have reasonably good classification. Therefore, even if our application falls in the second case, the MAP estimator is still a sensible choice.

To apply this theorem to our model, note that $\boldsymbol{W}$ corresponds to a composition of the edge formation probability $\Phi(\boldsymbol{\gamma}'\boldsymbol{\beta})$, the conditional weight distribution $\boldsymbol{h}$, and the scaling factor $\frac{\log(n)}{n}$. In the Monte Carlo exercises, and the real data application, we will calculate the CH divergence according to Definition 3 and (2.22)

$$W(k,l;q) = \frac{n}{\log(n)}\Phi(\ \overline{\boldsymbol{\gamma}(k,l)\prime\boldsymbol{\beta}}\ )h_{kl}(q) \tag{2.22}$$

where $\overline{\boldsymbol{\gamma}(k,l)\prime\boldsymbol{\beta}}$ is the median of $(\boldsymbol{\gamma}(k,l,z_i,z_j)\prime\boldsymbol{\beta})$ over (i,j) pairs such that $x_i = k, x_j = l$ or $x_i = l, x_j = k$. This accommodates our introduction of covariates in the edge formation process.

Estimation Algorithm

Apart from the theoretical challenge, the large size of the campaign finance network poses additional computational challenges. The parameter space of $\mathbf{x}$ is m^n. With 3 categories and 5,806 vertices, the parameter space is far larger than the number of atoms in the universe.[67] Therefore, exact solution to MAP in (2.14) is infeasible, and instead we need an efficient approximation method. In light of these considerations, we propose a Bayesian algorithm to approximate the posterior distribution of the latent ideology as well as other parameters.

In this Bayesian approach, the latent ideology vector $\mathbf{X}$, and the parameters $\epsilon, \boldsymbol{\theta}, \boldsymbol{\beta}$ are treated as random variables with certain prior probability distributions. Adjacency matrix $\mathbf{y}$ and reported affiliation $\hat{\mathbf{x}}$ are treated as one realization of the random variables $\mathbf{Y}$ and $\hat{\mathbf{X}}$. Observable characteristics $\mathbf{z}$ are treated as fixed and exogenous.

The prior distribution of the latent $\mathbf{X}$ is given by (2.4) in the network formation model. The prior distribution of $\boldsymbol{\theta}$, the parameter governing the unconditional

[67] According to Jackson (2010), the estimated number of atoms in the universe is on the order of 2^{270}.

probability distribution of ideology, is assumed to be a Dirichlet distribution:

$$\boldsymbol{\theta} \sim \mathrm{Dir}(\boldsymbol{\alpha}^{\theta}), \tag{2.23}$$

where $\boldsymbol{\alpha}^{\theta} \in \mathbb{R}_{+}^{m}$ is a vector of pre-specified concentration parameters. The prior distribution of ϵ, the parameter governing measurement error, is assumed to be a Beta distribution:

$$\epsilon \sim \mathrm{Beta}(\alpha_1^{\epsilon}, \alpha_2^{\epsilon}), \tag{2.24}$$

where $(\alpha_1^{\epsilon}, \alpha_2^{\epsilon}) \in \mathbb{R}_{+}^{2}$ is a vector of pre-specified concentration parameters. The prior distributions of $\mathbf{h} = \{\mathbf{h}_{kl}\}_{1 \le k \le l \le m}$, the conditional distribution of edge weight (transfer amount), is assumed to be a Dirichlet distribution:

$$\mathbf{h}_{kl} \sim \mathrm{Dir}(\boldsymbol{\alpha}^{h_{kl}}), \tag{2.25}$$

where $\boldsymbol{\alpha}^{h_{kl}} \in \mathbb{R}_{+}^{Q}$ is a vector of pre-specified concentration parameters. The prior distribution of $\boldsymbol{\beta}$, the parameter governing edge formation probability, is assumed to be a multivariate normal distribution:

$$\boldsymbol{\beta} \sim \mathcal{N}(\mathbf{0}, \tau^2 \mathbf{I}), \tag{2.26}$$

where $\tau \in \mathbb{R}_{+}$ is pre-specified standard deviation.

Given the prior probability distributions of $\mathbf{X}$, $\boldsymbol{\theta}$, $\boldsymbol{\beta}$, ϵ, and $\mathbf{h}$, the goal of Bayesian estimation is to update the belief on their joint distribution using data $\mathbf{y}, \hat{\mathbf{x}}$ and $\mathbf{z}$, i.e., to compute the posterior distribution $\mathbb{P}(\boldsymbol{\theta}, \boldsymbol{\beta}, \epsilon, \mathbf{h}, \mathbf{X} | \mathbf{y}, \hat{\mathbf{x}}, \mathbf{z})$. Once the posterior distribution is computed, it is straightforward to assess different objects of interest, especially the marginal distributions $\mathbb{P}(X_i | \mathbf{y}, \hat{\mathbf{x}}, \mathbf{z})$. There is neither an analytically nor a numerically convenient form to directly characterize of the joint posterior distribution. Fortunately, calculations of conditional distributions $\mathbb{P}(X_i|\mathbf{x_{-i}}, \mathbf{y}, \hat{\mathbf{x}}, \mathbf{z}, \boldsymbol{\theta}, \boldsymbol{\beta}, \epsilon, \mathbf{h},)$, $\mathbb{P}(\boldsymbol{\theta}|\mathbf{y}, \hat{\mathbf{x}}, \mathbf{z}, \boldsymbol{\beta}, \epsilon, \mathbf{h}, \mathbf{x})$, $\mathbb{P}(\boldsymbol{\beta}|\mathbf{y}, \hat{\mathbf{x}}, \mathbf{z}, \boldsymbol{\theta}, \epsilon, \mathbf{h}, \mathbf{x})$, $\mathbb{P}(\epsilon|\mathbf{y}, \hat{\mathbf{x}}, \mathbf{z}, \boldsymbol{\theta}, \boldsymbol{\beta}, \mathbf{h}, \mathbf{x})$, and $\mathbb{P}(\mathbf{h}|\mathbf{y}, \hat{\mathbf{x}}, \mathbf{z}, \boldsymbol{\theta}, \boldsymbol{\beta}, \epsilon, \mathbf{x})$ are relatively easy. Therefore, a Gibbs sampler algorithm is used to construct the joint posterior distribution. Gibbs sampler is a Markov Chain Monte Carlo (MCMC) algorithm, which repeatedly samples a set of random variables conditional on the values of all other random variables. It is particularly useful when sampling from conditional distributions is convenient.

Computing the Posterior Distribution $\mathbb{P}(\boldsymbol{\theta}, \boldsymbol{\beta}, \epsilon, \mathbf{h}, \mathbf{X}|\mathbf{y}, \hat{\mathbf{x}}, \mathbf{z})$. By Bayes' rule, the posterior distribution of X_i is given by:

$$\mathbb{P}(X_i = k|\mathbf{x_{-i}}, \mathbf{y}, \hat{\mathbf{x}}, \mathbf{z}; \boldsymbol{\theta}, \boldsymbol{\beta}, \epsilon, \mathbf{h},) = \frac{\mathbb{P}(\mathbf{y}, x_1, ..x_{i-1}, x_i = k, x_{i+1}, ...x_n, \hat{\mathbf{x}}|\mathbf{z}, \boldsymbol{\theta}, \boldsymbol{\beta}, \epsilon, \mathbf{h})}{\sum_{l=1}^{m} \mathbb{P}(\mathbf{y}, x_1, ..x_{i-1}, x_i = l, x_{i+1}, ...x_n, \hat{\mathbf{x}}|\mathbf{z}, \boldsymbol{\theta}, \boldsymbol{\beta}, \epsilon, \mathbf{h})},$$

which can be reduced to:

$$\mathbb{P}(X_i = k|\mathbf{x_{-i}}, \mathbf{y}, \hat{\mathbf{x}}, \mathbf{z}, \boldsymbol{\theta}, \boldsymbol{\beta}, \epsilon, \mathbf{h},) \propto$$

$$\begin{cases} \theta_k(1-\epsilon) \prod_{j \neq i} [\eta_{ij}(k, x_j) h_{kx_j}(y_{ij})]^{\mathbb{1}(y_{ij}>0)} [1 - \eta_{ij}(k, x_j)]^{\mathbb{1}(y_{ij}=0)} \text{ if } k = \hat{x}_i \\ \theta_k \dfrac{\epsilon}{m-1} \prod_{j \neq i} [\eta_{ij}(k, x_j) h_{kx_j}(y_{ij})]^{\mathbb{1}(y_{ij}>0)} [1 - \eta_{ij}(k, x_j)]^{\mathbb{1}(y_{ij}=0)} \text{ if } k \neq \hat{x}_i \end{cases} \quad \forall i \in \mathcal{V}^o, \tag{2.27}$$

where θ_k is the ideology prior, $(1-\epsilon)$ and $\frac{\epsilon}{m-1}$ are measurement accuracy of the report, and $\prod_{j \neq i} [\eta_{ij}(k, x_j) h_{kx_j}(y_{ij})]^{\mathbb{1}(y_{ij}>0)} [1 - \eta_{ij}(k, x_j)]^{\mathbb{1}(y_{ij}=0)}$ is information embedded in network connections. The posterior is an interaction of the three. When $\epsilon > 0$, i.e. allowing for measurement error, if the network data highly favors an ideology different from $\hat{x}_i$, it is possible for the posterior to override the prior, i.e. a posterior mode at $k \neq \hat{x}_i$. This can be viewed as a data-oriented correction of measurement error.

Similarly,

$$\mathbb{P}(X_i = k|\mathbf{x_{-i}}, \mathbf{y}, \hat{\mathbf{x}}, \mathbf{z}, \boldsymbol{\theta}, \boldsymbol{\beta}, \epsilon, \mathbf{h},) \propto \\ \theta_k \prod_{j \neq i} [\eta_{ij}(k, x_j) h_{kx_j}(y_{ij})]^{\mathbb{1}(y_{ij}>0)} [1 - \eta_{ij}(k, x_j)]^{\mathbb{1}(y_{ij}=0)} \qquad \forall i \in \mathcal{V} \backslash \mathcal{V}^o, \tag{2.28}$$

where θ_k is the ideology prior, and $\prod_{j \neq i} [\eta_{ij}(k, x_j) h_{kx_j}(y_{ij})]^{\mathbb{1}(y_{ij}>0)} [1 - \eta_{ij}(k, x_j)]^{\mathbb{1}(y_{ij}=0)}$ is information embedded in network connections. The posterior is an interaction of the two. Summoning conjugacy, the posterior distribution of $\boldsymbol{\theta}$ is given by:

$$\boldsymbol{\theta}|\mathbf{y}, \hat{\mathbf{x}}, \mathbf{z}, \boldsymbol{\beta}, \epsilon, \mathbf{h}, \mathbf{x} \sim \text{Dir}(\boldsymbol{\alpha}^\theta + [n_k]_{1 \leq k \leq m}), \tag{2.29}$$

where $n_k = \sum_{1 \leq i \leq n} \mathbb{1}(x_i = k)$ is the number of vertices with x_i equal to k. The posterior distribution of ϵ is given by:

$$\epsilon|\mathbf{y}, \hat{\mathbf{x}}, \mathbf{z}, \boldsymbol{\theta}, \boldsymbol{\beta}, \mathbf{h}, \mathbf{x} \sim \text{Beta}(\alpha_1^\epsilon + n_e, \alpha_2^\epsilon + n_t), \tag{2.30}$$

where $n_e = \sum_{i \in \mathcal{V}^o} \mathbb{1}(\hat{x}_i \neq x_i)$ is the number of vertices whose self report is different from its ideology, and $n_t = \sum_{i \in \mathcal{V}^o} \mathbb{1}(\hat{x}_i = x_i)$ is the number of vertices whose self report is the same as its ideology. The posterior distribution of $\mathbf{h}$ is given by:

$$\mathbf{h}_{kl}|\mathbf{y}, \hat{\mathbf{x}}, \mathbf{z}, \boldsymbol{\theta}, \boldsymbol{\beta}, \epsilon, \mathbf{x} \sim \text{Dir}(\boldsymbol{\alpha}^{h_{kl}} + [n_{kl,q}]_{1 \leq q \leq Q}), \tag{2.31}$$

where $n_{kl,q} = \sum_{1 \leq i < j \leq n} \max\{\mathbb{1}(x_i = k, x_j = l), \mathbb{1}(x_i = l, x_j = k)\} \mathbb{1}(y_{ij} = q)$ is the number of edges with transfer amount w_q between PCs of ideologies k and l. Constructing the posterior distribution of $\boldsymbol{\beta}$ is more delicate. $\boldsymbol{\beta}$ is essentially the coefficient vector in a Probit regression model, whose posterior distribution does not have an analytically convenient form. Therefore, instead of directly sampling from a closed form distribution, a data augmentation strategy introduced in Albert and Chib

(1993) is used. First, sample auxiliary variable $\mathbf{u} = \{u_{ij}\}_{1\le i<j\le n}$ from the following truncated normal distributions:

$$u_{ij} \sim \begin{cases} \mathcal{N}(\boldsymbol{\gamma}(x_i, x_j, \mathbf{z}_i, \mathbf{z}_j)'\boldsymbol{\beta}, 1)|u_{ij} > 0 \text{ if } y_{ij} > 0 \\ \mathcal{N}(\boldsymbol{\gamma}(x_i, x_j, \mathbf{z}_i, \mathbf{z}_j)'\boldsymbol{\beta}, 1)|u_{ij} < 0 \text{ if } y_{ij} = 0. \end{cases} \tag{2.32}$$

Conditional on the auxiliary variable $\mathbf{u}$, the posterior distribution of $\boldsymbol{\beta}$ is given by:

$$\boldsymbol{\beta}|\mathbf{y}, \hat{\mathbf{x}}, \mathbf{z}, \boldsymbol{\theta}, \epsilon, \mathbf{h}, \mathbf{x}, \mathbf{u} \sim \mathcal{N}\left((\tau^{-2}\mathbf{I} + \boldsymbol{\Gamma}'\boldsymbol{\Gamma})^{-1}\boldsymbol{\Gamma}'\mathbf{u},\ (\tau^{-2}\mathbf{I} + \boldsymbol{\Gamma}'\boldsymbol{\Gamma})^{-1}\right), \tag{2.33}$$

where $\boldsymbol{\Gamma} = \left[\boldsymbol{\gamma}(x_i, x_j, \mathbf{z}_i, \mathbf{z}_j)'\right]_{1\le i<j\le n}$.

The Gibbs sampler algorithm is summarized below:

1. Initialize $\mathbf{x}^0, \boldsymbol{\theta}^0, \boldsymbol{\beta}^0,\ \epsilon^0,\ \mathbf{h}^0$.

2. Iteratively sample from conditional posterior distribution. Specifically, in iteration t, we sample one set of parameters $(\mathbf{x}^t, \boldsymbol{\theta}^t, \epsilon^t, \mathbf{h}^t, \boldsymbol{\beta}^t)$ with the following procedure:

   a) Sample $\{x_i\}^t_{1\le i\le n}$ sequentially from distribution $\mathbb{P}(X_i|\mathbf{y}, \hat{\mathbf{x}}, \mathbf{z}, \boldsymbol{\theta}^{t-1}, \boldsymbol{\beta}^{t-1}, \epsilon^{t-1}, \mathbf{h}^{t-1}, x^t_1, ..x^t_{i-1}, x^{t-1}_{i+1}, ...x^{t-1}_n)$ using (2.27) and (2.28).

   b) Sample vector $\boldsymbol{\theta}^t$ from distribution $\mathbb{P}(\boldsymbol{\theta}|\mathbf{y}, \hat{\mathbf{x}}, \mathbf{z}, \boldsymbol{\beta}^{t-1}, \epsilon^{t-1}, \mathbf{h}^{t-1}, \mathbf{x}^t)$ using (2.29).

   c) Sample ϵ^t from distribution $\mathbb{P}(\epsilon|\mathbf{y}, \hat{\mathbf{x}}, \mathbf{z}, \boldsymbol{\beta}^{t-1}, \mathbf{h}^{t-1}, \mathbf{x}^t, \boldsymbol{\theta}^t)$ using (2.30).

   d) Sample vectors $\mathbf{h}^t$ from distribution $\mathbb{P}(\mathbf{h}|\mathbf{y}, \hat{\mathbf{x}}, \mathbf{z}, \boldsymbol{\beta}^{t-1}, \mathbf{x}^t, \boldsymbol{\theta}^t, \epsilon^t)$ using (2.31)

   e) Sample auxiliary vector $\mathbf{u}^t$ from distribution $\mathbb{P}(\mathbf{u}|\mathbf{y}, \mathbf{z}, \boldsymbol{\beta}^{t-1}, \mathbf{x}^t)$ using (2.32).

   f) Sample vector $\boldsymbol{\beta}^t$ from distribution $\mathbb{P}(\boldsymbol{\beta}|\mathbf{y}, \hat{\mathbf{x}}, \mathbf{z}, \mathbf{x}^t, \boldsymbol{\theta}^t, \epsilon^t, \mathbf{h}^t, \mathbf{u}^t)$ using (2.33).

3. Burn in the first T_1 iterations. Use samples from iterations $T_1 + 1$ to $T_1 + T_2$ to construct posterior distribution.

The first step initializes the Markov chain. The initial values should not affect the steady state and can be determined either at random or by other algorithms. We use the former in our application. The second step simulates the Markov chain by repeatedly sampling one parameter conditional on the values of all other parameters. The sampling order of the parameters is arbitrary, and a different order can be used, e.g. one can sample $\boldsymbol{\theta}$ before $\mathbf{x}$. In order to speed up convergence, we use the newly sampled parameter immediately in the following sampling procedures and do not wait until the next iteration, e.g., $\mathbf{x}^t$ sampled from (a) is used in the sampling of $\boldsymbol{\theta}^t$ in (b). The final step discards the initial portion of the Markov chain, namely the first T_1 iterations, where steady state is not reached. Pooling the remainder samples gives an approximate joint posterior distribution $\mathbb{P}(\boldsymbol{\theta}, \boldsymbol{\beta}, \epsilon, \mathbf{h}, \mathbf{X} | \mathbf{y}, \hat{\mathbf{x}}, \mathbf{z})$.

2.5 Monte Carlo Evidence

In this section, we present Monte Carlo evidence to evaluate the performance of the community recovery algorithm proposed in Section 2.4. We conduct four sets of Monte Carlo simulations that differ in the specifications of the edge formation process and the network size, both of which affect the Chernoff-Hellinger divergence measure as defined in (2.18) because they enter the expression of $W(k, l; q)$ (see Eq. (2.22)).

The first three sets of Monte Carlo simulations share a framework that is similar to the homogeneous labeled stochastic block model as in Jog and Loh (2015), even though the edge distributions are not exactly homogeneous due to the introduction of the covariates in the edge formation probability. Our Monte Carlo results show a strong confirmation of Theorem 1 that CH divergence of 1 is a sharp threshold for exact recovery. In the first specification, the network edge formation patterns and the network size imply a CH divergence of 1.0074, and we find that the misclassification rate is on average 1.10%. The second specification differs from the first one only in the weight distribution, resulting in a smaller CH divergence of 0.4719 and we find a larger average misclassification rate of 5.68%. The third specification differs from the second one only in the network size, which results in a larger CH divergence of 1.7537 and we find a smaller average misclassification rate of 0. In the fourth set of Monte Carlo simulations, the network size is similar to the real data (about 6,000 vertices), and the edge distributions are heterogeneous in a flexible way, resulting in a CH divergence of 6.0110. This is intended to assess the performance of the algorithm in a data set that resembles the real data. The results show that the algorithm performs surprisingly well with an average misclassification rate of 0.0002%, although it is slower due to the scale of the network. We summarize our main simulation results in the text, but many of the less essential details are left in Appendices 4.3-4.6.

Common Specifications Across All Four Sets of Monte Carlo Simulations. The specifications that are common across all four sets of Monte Carlo studies are listed as follows. The aspects of the specifications that are unique to each set of Monte Carlo studies, as well as their specific results, are presented separately in subsequent subsections.

- The number of ideologies: $m = 3$.

- Marginal distribution of ideology: $\theta = (1/3, 1/3, 1/3)$.

- Probability of measurement error: $\epsilon = 0.05$.
- Fraction of vertices with reported affiliation: 40%.
- Edge formation probability is specified by $y_{ij} > 0$ iff

$$\boldsymbol{\gamma}(x_i, x_j, \mathbf{z}_i, \mathbf{z}_j)'\boldsymbol{\beta} + e_{ij} > 0,$$

where $\boldsymbol{\gamma}(x_i, x_j, \mathbf{z}_i, \mathbf{z}_j) = (\boldsymbol{\gamma_x}, \boldsymbol{\gamma_z})$ is a list of 19 variables and $\boldsymbol{\beta} = (\boldsymbol{\beta_x}, \boldsymbol{\beta_z}) \in \mathbb{R}^{19}$. To make the simulated data as close as possible to the real data we use for the empirical analysis, we include in the vector $\mathbf{z}_i$ the PC's budget, state, industry, dummy for House campaign, dummy for Senate campaign, dummy for Presidential campaign, dummy for qualified PAC, dummy for qualified Party, dummy for national committee, dummy for authorized by a candidate, and dummy for joint fund-raiser, and we construct $\boldsymbol{\gamma_z} \in \mathbb{R}^{13}$ based on $(\mathbf{z}_i, \mathbf{z}_j)$; specifically,

$$\boldsymbol{\gamma_z} = \begin{pmatrix} \mathbb{1}_{\text{state}_i = \text{state}_j}, \mathbb{1}_{\text{industry}_i = \text{industry}_j}, \\ \mathbb{1}_{(\text{house}_i=1)\vee(\text{house}_j=1)}, \mathbb{1}_{(\text{senate}_i=1)\vee(\text{senate}_j=1)}, \mathbb{1}_{(\text{president}_i=1)\vee(\text{president}_j=1)}, \\ \mathbb{1}_{(\text{qualified PAC}_i=1)\vee(\text{qualified PAC}_j=1)}, \mathbb{1}_{(\text{qualified Party}_i=1)\vee(\text{qualified Party}_j=1)}, \\ \mathbb{1}_{(\text{national}_i=1)\vee(\text{national}_j=1)}, \mathbb{1}_{(\text{authorized}_i=1)\vee(\text{authorized}_j=1)}, \mathbb{1}_{(\text{fundraiser}_i=1)\vee(\text{fundraiser}_j=1)}, \\ [\ln b_i + \ln b_j], [(\ln b_i)^2 + (\ln b_j)^2], \ln b_i \ln b_j \end{pmatrix}. \tag{2.34}$$

We set

$$\beta_{\mathbf{z}} = \begin{pmatrix} 0.3, 0.3, 0.1, 0.1, 0.1, 0.2, 0.2, \\ 0.15, 0.15, 0.15, 0.01, -0.01, 0.001 \end{pmatrix}$$

in all four sets of Monte Carlo simulations.

$\boldsymbol{\gamma}_{\boldsymbol{x}} \in \mathbb{R}^6$ includes a constant term and interaction terms of x_i and x_j; specifically,

$$\boldsymbol{\gamma}_{\boldsymbol{x}} = \begin{pmatrix} 1, \mathbb{1}_{x_i=x_j=\text{Dem}},\ \mathbb{1}_{x_i=x_j=\text{Rep}},\ \mathbb{1}_{x_i=x_j=\text{Ind}}, \\ \mathbb{1}_{(x_i=\text{Dem},x_j=\text{Ind})\vee(x_i=\text{Ind},x_j=\text{Dem})},\ \mathbb{1}_{(x_i=\text{Rep},x_j=\text{Ind})\vee(x_i=\text{Ind},x_j=\text{Rep})} \end{pmatrix}. \tag{2.35}$$

- Transfer amount is discretized into four bins. Therefore, conditional on $y_{ij} > 0$, $y_{ij} \in \{1, 2, 3, 4\}$.

Monte Carlo I: 500 Networks with $n = 100$, CH Divergence Exceeding 1

In the first set of Monte Carlo simulations, 500 networks are simulated and estimated. They have network size $n = 100$, the coefficients in the network formation probability are given by $\boldsymbol{\beta}_{\boldsymbol{x}} = (-1.5, 0.5, 0.5, 0.5, 0, 0)$, and the edge's weight distributions are given by $\mathbf{h}_{\text{Dem,Dem}} = \mathbf{h}_{\text{Rep,Rep}} = \mathbf{h}_{\text{Ind,Ind}} = (0.05, 0.1, 0.4, 0.45)$, and $\mathbf{h}_{\text{Dem,Rep}} = \mathbf{h}_{\text{Dem,Ind}} = \mathbf{h}_{\text{Rep,Ind}} = (0.4, 0.3, 0.2, 0.1)$. Note that the specification of $\boldsymbol{\beta}_{\boldsymbol{x}} = (-1.5, 0.5, 0.5, 0.5, 0, 0)$ implies that the link formation depends on whether the two vertices are of the same ideology or are of differen ideologies, and fits into the homogeneity special case we described in Eq. (2.20). The implied CH divergence according to (2.21) and (2.22) is $1.0074 > 1$.

For these simulations, the total execution time is 235,405 seconds (about 65 hours). The speed of convergence in terms of the number of iterations varies, which is shown in Figure 4.1 where we plot the histogram of the number of iterations (including burn-in and posterior) across the 500 networks we simulated. The distribution of misclassification rates is summarized in Figure 4.2 and Table 4.3.[68] They are small

[68] Estimated ideology is defined as the posterior mode.

in general, and in 37.4% of the simulations, there is no misclassification.[69] Moreover, 98.8% of the simulations have misclassification rates lower than the measurement error rate 0.05, which implies that in most cases our algorithm successfully corrects some of the misreports. Table 4.4 provides a detailed tabulation of the estimated vs. true ideologies.

The results above focus only on the posterior mode, and the following analysis further investigates the patterns of the posterior distributions. For the correctly classified vertices, our categorical classification based on posterior mode is rather precise. Figure 4.3 shows that the differences in posterior probability between the highest posterior probability (i.e., the posterior of the true ideology) and the second highest posterior probability are highly concentrated around 1. This indicates that for most of these vertices, the posterior distribution is strongly informative of the true ideology. For the misclassified vertices, however, the scales of misclassification vary. Figure 4.4 shows that the differences in the posterior probability between the highest posterior probability and the posterior probability of the true ideology are approximately uniformly distributed between 0 where the classification is only a bit off and 1 where the classification is far off. This suggests that the misclassification is likely to be caused by unusual realizations of the networks process rather than the failure of our estimation algorithm. The randomness in the network formation renders the ideology information unclear or even misleading for some vertices, though this occurs rarely. Additional analysis of the posterior mean of other parameters are

[69] 37.4% exact recovery is lower than the theoretical prediction, for several reasons. First, probability of exact recovery approaches 1 only when network size approaches infinity. In this case, the network size is only 100, which may be too small. Second, our model is not exactly the same as the model in Jog and Loh (2015) because we include covariates in the edge formation probability, which is more complicated. Third, we do not use an exact MAP estimator, which may also introduce approximation error.

presented in Tables 4.5-4.8. These tables show that the algorithm recovers the true parameters effectively, except for the weight distribution parameters $\mathbf{h}_{kl}$. We will show that these parameters will be estimated more precisely as the network size gets larger in Monte Carlo simulations III (where $n = 500$) and IV (where $n = 6000$).

Monte Carlo II: 500 Networks with $n = 100$, CH Divergence Less Than 1

The specifications for the second set of Monte Carlo simulations are the same as those for the first set except for the weight distributions $\mathbf{h}_{\text{Dem,Dem}}$, $\mathbf{h}_{\text{Rep,Rep}}$ and $\mathbf{h}_{\text{Ind,Ind}}$. Specifically, the edge's weight distributions are given by $\mathbf{h}_{\text{Dem,Dem}} = \mathbf{h}_{\text{Rep,Rep}} = \mathbf{h}_{\text{Ind,Ind}} = (0.2, 0.15, 0.35, 0.3)$, and $\mathbf{h}_{\text{Dem,Rep}} = \mathbf{h}_{\text{Dem,Ind}} = \mathbf{h}_{\text{Rep,Ind}} = (0.4, 0.3, 0.2, 0.1)$. Again, 500 networks with size $n = 100$ are simulated and estimated. The implied CH divergence according to (2.21) and (2.22) decreases to 0.4719, which is now less than 1.

The total execution time for these simulations is 418,322 seconds (about 5 days). The speed of convergence in terms of the number of iterations is relatively slow; Figure 4.5 depicts the histogram of the number of iterations (including burn-in and posterior). Figure 4.6 and Table 4.9 summarize the distribution of misclassification rates across the 500 simulations. Comparing with Monte Carlo I, the misclassification rates are higher; for the worst case, the misclassification rate is as high as 24%. Only 2.8% of the simulations have 0 misclassification. This is consistent with the theoretical prediction of Theorem 1 that a CH divergence lower than 1 is associated with low probability of exact recovery. Table 4.10 provides a detailed tabulation of the estimated vs. true ideologies in this set of Monte Carlo simulations. Figure 4.7 plots, for the correctly classified vertices, the histogram of the difference in posterior probability between the highest posterior probability (i.e., the posterior for the true ideology) and the second highest posterior probability. It is shown to be mostly concentrated at 1, though relative to Figure 4.3 for Monte Carlo I, the difference is somewhat more likely to be less than 1 and is more spread out. This indicates that, when CH is less than 1, our categorizations are not as informative even though we obtained the correct classification. Similarly, Figure 4.8 plots, for the misclassified vertices, the histogram of the difference in posterior probability between the highest posterior probability and the posterior probability of the true ideology. The difference is rather evenly distributed, which is similar to Figure 4.4 in Monte Carlo I. Additional analysis of the posterior mean of other parameters are presented in Tables 4.11-4.14. These tables show that the algorithm recovers the true parameters effectively, except for the weight distribution parameters $\mathbf{h}_{kl}$.

Monte Carlo III: 500 Networks with $n = 500$, CH Divergence Exceeding 1

The specifications for the third set of Monte Carlo simulations are the same as those of Monte Carlo II except for the network size. Now we simulate 500 networks, each with a network size of $n = 500$. As a result of the increase in the network size, the implied CH divergence according to (2.21) and (2.22) is now 1.7537, which is larger than 1.

The total execution time for these simulations is 31,974 seconds (about 9 hours). The speed of convergence is fast in terms of the number of iterations (ranging from 600 to 900); Figure 4.9 depicts the histogram of the number of iterations (including burn-in and posterior). The misclassification rates in all 500 simulations are 0. Therefore, this set of simulations over-perform the previous two in terms of both convergence speed and accuracy rate. Figure 4.10 plots the histogram of the difference in posterior probability between the highest posterior probability (i.e., the posterior for the true ideology) and the second highest posterior probability for the correctly classified vertices (which are all vertices because of 0 misclassification rate). The difference is almost completely concentrated at 1, indicating that our categorization based on posterior mode is very informative; in fact, for 99.9976% of the correctly classified vertices, the posterior probability on the true ideology is 1. Additional analysis of the posterior mean of other parameters are presented in Tables 4.15-4.18. These tables show that the algorithm recovers the true parameters effectively, including the weight distribution parameters $\mathbf{h}_{kl}$ (see Table 4.17).

Monte Carlo IV: 100 Networks with $n = 6,000$, CH Divergence Exceeding 1

In the fourth set of Monte Carlo simulations, we make several important changes. First, we increase the network size to $n = 6000$, which is comparable to the size of the political contribution network in our data. Also importantly, we deviate from the symmetry in the within-community and between-community link formation probabilities. Both changes are intended to assess the performance of our estimation algorithm in an environment that resembles the actual data. We again simulate and estimate 100 networks. Specifically, the coefficient in the network formation probability is given by $\boldsymbol{\beta_x} = (-3, 1, 1, 0.7, 0.3, 0.3)$, and the edge's weight distributions are given by $\mathbf{h}_{\text{Dem,Dem}} = \mathbf{h}_{\text{Rep,Rep}} = (0.1, 0.2, 0.2, 0.5)$, $\mathbf{h}_{\text{Ind,Ind}} = (0.25, 0.25, 0.25, 0.25)$, $\mathbf{h}_{\text{Dem,Rep}} =$

$(0.5, 0.2, 0.2, 0.1)$, and $\mathbf{h}_{\text{Dem,Ind}} = \mathbf{h}_{\text{Rep,Ind}} = (0.3, 0.3, 0.3, 0.1)$. Using expressions (2.18) and (2.22) to evaluate pairwise divergence between ideology communities, we have $D_{L+}(\boldsymbol{\theta}, \boldsymbol{W}(\text{Dem}), \boldsymbol{W}(\text{Rep})) = 13.1003$, and $D_{L+}(\boldsymbol{\theta}, \boldsymbol{W}(\text{Dem}), \boldsymbol{W}(\text{Ind})) = D_{L+}(\boldsymbol{\theta}, \boldsymbol{W}(\text{Rep}), \boldsymbol{W}(\text{Ind})) = 6.011$, and thus the CH divergence $6.011 > 1$.

The total execution time for the fourth set of simulations is 930,673 seconds (about 11 days). The speed of convergence is fast in terms of the number of iterations (ranging from 600 to 900); Figure 4.11 depicts the histogram of the number of iterations. Therefore, the long execution time is a result of heavy computation in each iteration, not the number of iteration. Misclassification rates are 0 for 99 simulations, 0.0167% (i.e., 1 vertex is misclassified) for 1 simulation. For the correctly classified vertices, the numerical posterior distributions of ideology are degenerate in the true ideology. For the only vertex that is incorrectly classified in this set of simulations, the difference between the highest posterior probability and the posterior probability of the true ideology is 0.13. These results suggest that the simulated data exhibit strong information on the community structure, and that our algorithm is efficient in identifying this structure. Additional analysis of the posterior mean of other parameters are presented in Tables 4.19-4.22. These tables show that the algorithm recovers the true parameters effectively, including the weight distribution parameters $\mathbf{h}_{kl}$.

To summarize, our algorithm has excellent performance when the data is generated with CH divergence greater than 1. It has reasonably good performance even when the data is generated with CH divergence lower than 1. The results also suggest that a large network is not necessarily undesirable. On the one hand, it brings in more computational burden and increases the runtime; on the other hand, it also embodies more information and speeds up the convergence.

2.6 Empirical Implementation and Results

Empirical Implementation.

We empirically infer the ideologies of 5,806 PCs from the giant component in the campaign finance network depicted in Figure 2.3. There are 3 categories of ideologies: Democratic, Republican, and Independent. For the small number of PCs whose self-reported affiliations do not belong to these categories, we treat them as if we do not observe their report. The set $\mathcal{V}^o$ contains PCs with self-reported affiliations $\hat{x}_i$'s. We assume the following functional form for the edge formation probability:

$$\begin{aligned}
&\boldsymbol{\gamma}(x_i, x_j, \mathbf{z}_i, \mathbf{z}_j)'\boldsymbol{\beta} \\
&= \beta_1 + \beta_2 \mathbb{1}_{x_i=x_j=\text{Dem}} + \beta_3 \mathbb{1}_{x_i=x_j=\text{Rep}} + \beta_4 \mathbb{1}_{x_i=x_j=\text{Ind}} \\
&+ \beta_5 \mathbb{1}_{(x_i=\text{Dem},x_j=\text{Ind})\vee(x_i=\text{Ind},x_j=\text{Dem})} + \beta_6 \mathbb{1}_{(x_i=\text{Rep},x_j=\text{Ind})\vee(x_i=\text{Ind},x_j=\text{Rep})} \\
&+ \beta_7 \mathbb{1}_{\text{state}_i=\text{state}_j} + \beta_8 \mathbb{1}_{\text{industry}_i=\text{industry}_j} + \beta_9 \mathbb{1}_{(\text{house}_i=1)\vee(\text{house}_j=1)} \\
&+ \beta_{10} \mathbb{1}_{(\text{senate}_i=1)\vee(\text{senate}_j=1)} + \beta_{11} \mathbb{1}_{(\text{president}_i=1)\vee(\text{president}_j=1)} \\
&+ \beta_{12} \mathbb{1}_{(\text{qualified PAC}_i=1)\vee(\text{qualified PAC}_j=1)} + \beta_{13} \mathbb{1}_{(\text{qualified Party}_i=1)\vee(\text{qualified Party}_j=1)} \\
&+ \beta_{14} \mathbb{1}_{(\text{national}_i=1)\vee(\text{national}_j=1)} + \beta_{15} \mathbb{1}_{(\text{authorized}_i=1)\vee(\text{authorized}_j=1)} \\
&+ \beta_{16} \mathbb{1}_{(\text{fundraiser}_i=1)\vee(\text{fundraiser}_j=1)} + \beta_{17} \left[\ln b_i + \ln b_j\right] + \beta_{18} \left[(\ln b_i)^2 + (\ln b_j)^2\right] \\
&+ \beta_{19} \ln b_i \ln b_j,
\end{aligned} \tag{2.36}$$

where the first term is a constant characterizing the baseline connection probability between Democratic and Republican PCs; the second to the sixth terms characterize the connection probabilities for other ideology pairs; the seventh and the eighth terms capture the effect of the two PCs belonging to the same state or industry; the ninth to the sixteenth terms capture the effects of PCs' institutional characteristics: whether one of them is a House campaign, a Senate campaign, a Presidential campaign, a qualified PAC, a qualified Party, a national committee, authorized by a candidate, or a joint fundraiser; and the seventeenth to the nineteenth terms capture the effect of both PCs' budgets on link formations probability, which is a restrictive form of that in (2.8) and assumes that the effect of financial and institutional characteristics are the same across ideological pairs. The main reason for this parsimonious specification is to reduce the computational intensity. The estimation results do not seem to show signs of severe mis-specification. The transfer amount Y_{ij} is discretized into multiples of \$500, and can take values of $\{0, 1, 2, ..., 100\}$ where 100 includes all the transfer higher than \$50,000. The initial values in the Gibbs sampler are randomly generated, and different sets of initial values are used.

Estimation Results: Posterior Mean and Standard Deviations.

Table 2.10 presents the posterior mean and standard deviation of $\boldsymbol{\beta}$, the coefficients in edge formation probability. The second to the sixth coefficients are all positive, indicating that Democratic and Republican PCs (the baseline case) have the lowest connection probability. Additionally, the Democratic PCs have stronger within party connection than the Republican PCs. Moreover, Independent PCs have a higher probability of connecting with Democratic or Republican PCs than other Independent PCs. It is also interesting to note that, *everything else equal,* pairs of Republican PCs are less likely to form a link than Republican/Independent or Democratic/Independent pairs of PCs.

Tables 2.11 and 2.12 present the posterior mean and standard deviation of $\boldsymbol{\theta}$, the unconditional probability of ideology; and ϵ, the measurement error. Tables 2.11 shows that in the population of all PCs, 40.01% are Democratic, 42.74% are Republican, and 17.24% are Independent. The posterior standard deviations of these estimates are small. Table 2.12 shows that the self-reported ideologies of the PCs are likely to be erroneous with probability 7.37%.

Due to the large number of parameters in the weight distribution function, we present the estimates of $\mathbf{h}$ in Figure 2.10, which shows the posterior means of all the values of $(h_{kl,1}, ..., h_{kl,100})$ for all $k, l \in$ {Dem, Rep, Ind} pairs graphically.

$\boldsymbol{\beta}$	Posterior Mean	Posterior Standard Deviation
constant	-7.4729	0.0166
$\mathbb{1}_{x_i=x_j=Dem}$	1.3301	0.0099
$\mathbb{1}_{x_i=x_j=Rep}$	0.8210	0.0130
$\mathbb{1}_{x_i=x_j=Ind}$	0.6405	0.0134
$\mathbb{1}_{(x_i=Dem,x_j=Ind)\vee(x_i=Ind,x_j=Dem)}$	1.2909	0.0121
$\mathbb{1}_{(x_i=Rep,x_j=Ind)\vee(x_i=Ind,x_j=Rep)}$	1.5797	0.0119
Same state	0.6399	0.0043
Same industry	0.2185	0.0117
One of them is a House campaign	0.5306	0.0034
One of them is a Senate campaign	0.3783	0.0035
One of them is a Presidential campaign	0.0212	0.0113
One of them is a qualified PAC	0.7006	0.0049
One of them is a qualified Party	-0.5334	0.0066
One of them is a national committee	0.9421	0.0133
One of them is authorized by a candidate	-0.4473	0.0090
One of them is a joint fundraiser	-0.7623	0.0059
$(\ln b_i + \ln b_j)$	-0.0442	0.0023
$((\ln b_i)^2 + (\ln b_j)^2)$	0.0162	0.00004
$\ln b_i \ln b_j$	-0.0010	0.0002

Table 2.10: Posterior Distribution of β

$\boldsymbol{\theta}$	Posterior Mean	Posterior Standard Deviation
$\mathbb{P}$(Dem)	0.4001	0.0045
$\mathbb{P}$(Rep)	0.4274	0.0046
$\mathbb{P}$(Ind)	0.1724	0.0036

Table 2.11: Posterior Distribution of θ

$\boldsymbol{\epsilon}$	Posterior Mean	Posterior Standard Deviation
	0.0737	0.0042

Table 2.12: Posterior Distribution of ϵ

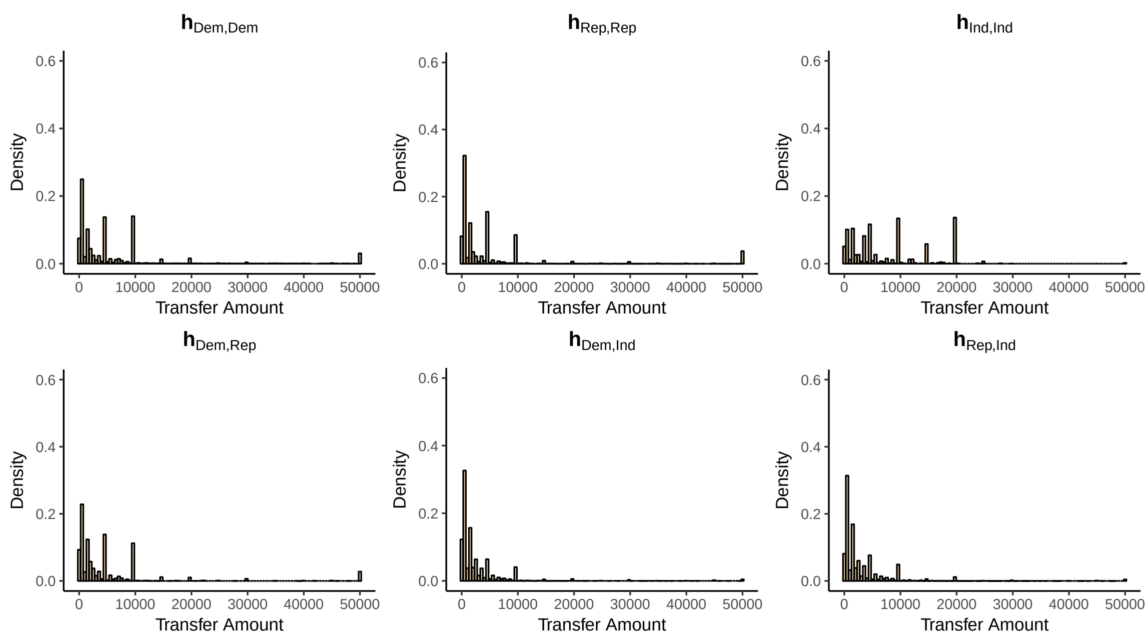

Figure 2.10: Posterior Mean of h

Note: Bin size is 500. Probability of transfer amount higher than \$50,000 is plotted at \$50,000.

Chernoff-Hellinger Divergence of the Estimated Model. Using the posterior mode of $\mathbf{x}$, and the posterior means of $\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{h}$, we calculate the implied Chernoff-Hellinger divergence:

$$D_{L+}(\boldsymbol{\theta}, \boldsymbol{W}(\text{Dem}), \boldsymbol{W}(\text{Rep})) = 14.7760,$$

$$D_{L+}(\boldsymbol{\theta}, \boldsymbol{W}(\text{Dem}), \boldsymbol{W}(\text{Ind})) = 26.6116,$$

$$D_{L+}(\boldsymbol{\theta}, \boldsymbol{W}(\text{Rep}), \boldsymbol{W}(\text{Ind})) = 22.4425,$$

and thus the CH divergence is $14.7760 > 1$. Therefore, the data generating process, corresponding to our estimated parameters, satisfies the condition for exact recovery as stated in Theorem 1.

Comparing Estimated and Self-Reported Ideologies for PCs that Self Report Ideologies

Using the posterior mode as a point estimate of the ideology, Table 2.13 presents the cross tabulation of all PCs according to self-reported and estimated ideology. Overall, 90.70% of our estimates match the self reports: 94.36% for self-reported Democratic PCs, and 89.49% for self-reported Republican PCs.

	Estimated Dem	Estimated Rep	Estimated Ind
Self-Reported Dem	954 (94.36%)	43 (4.25%)	14 (1.38%)
Self-Reported Rep	46 (4.60%)	894 (89.49%)	59 (5.91%)
Self-Reported Ind	16 (29.09%)	14 (25.45%)	25 (45.45%)
No Reported Affiliation	748 (20.00%)	1,202 (32.13%)	1,791 (47.87%)

Table 2.13: Tabulation of Estimated vs. Self-Reported Ideology
Note: The percentages are calculated for each row.

Re-examining the Network Statistics Using Estimated Ideologies

In Section 2.3, we presented the network statistics based on self-reported ideologies for those PCs that self reported their ideologies. Here we re-examine these statistics based on estimated ideologies of all PCs. Conditional on the estimated ideologies, the empirical distribution of transfer amount is shown in Figure 2.11, and the mean of each distribution is shown in Table 2.14. On average, the transfer amount is the highest for Democratic PC pairs and Republican PC pairs, and this is partially due to the heavy tail of the within-party contributions. The transfer amount is smaller between Democratic and Republican PCs, and smallest when it involves Independent PCs. This is consistent with the estimates on connection probability. Independent PCs have overall high connection probability, but the associated transfer amount is small. Democratic and Republican PCs have relatively lower within-party connection probability, but the associated transfer amount is larger.

Dem, Dem	Rep, Rep	Ind, Ind	Dem, Rep	Dem, Ind	Rep, Ind
26,311.03	22,165.53	8,144.17	18,507.13	3,206.26	3,769.65

Table 2.14: Mean of Transfer Amount Conditional on Estimated Ideology

Next, we compare the contribution patterns of the PCs according to their self-reported vs. estimated ideologies. For each PC i, let numDem_i, numRep_i, and numInd_i denote its numbers of connections with (estimated) Democratic, Republican, and Independent PCs respectively. In Figures 2.12-2.14, we plot the distributions of numDem (dark bar), numRep (white bar), and numInd (gray bar) for different groups of PCs, focusing on the difference between self-reported and estimated ideologies. The

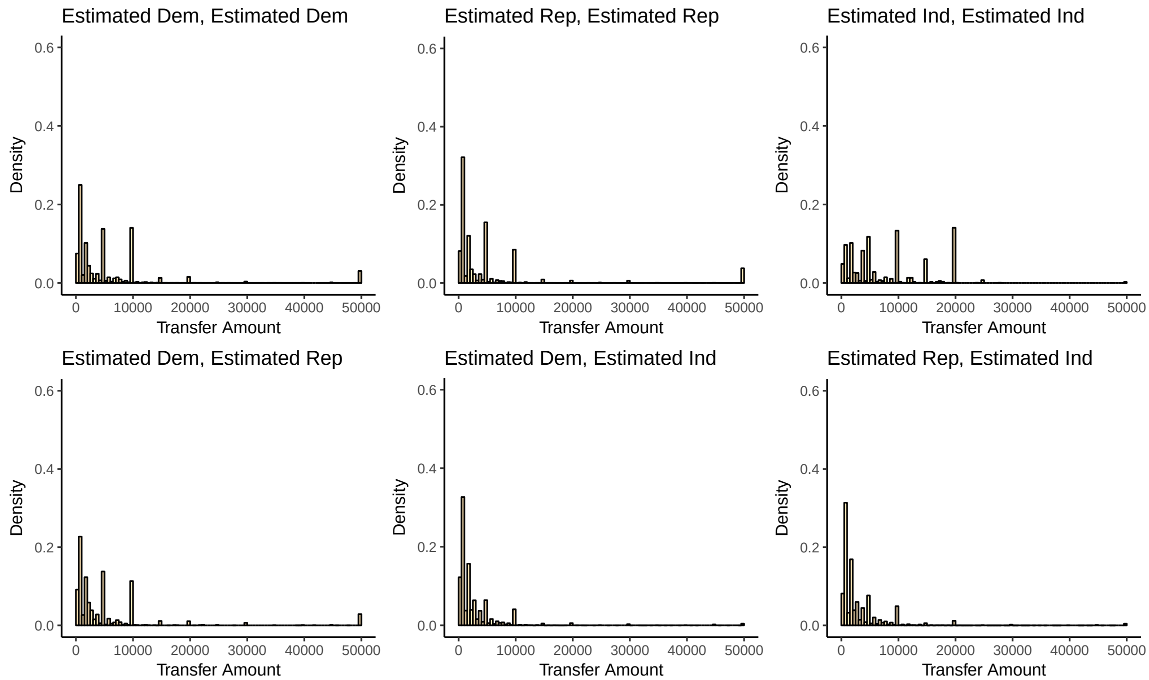

Figure 2.11: Empirical Distribution of Transfer Amount Conditional on Estimated Ideology

Note: Bin size is 500. Observations with transfer amount higher than $50,000 are plotted at $50,000.

left panel in Figure 2.12 presents the distributions for PCs that *self reported* to be Democratic, and the right panel for PCs that did not self report but are estimated to be Democratic PCs. Figures 2.13 and 2.14 are similar, but for Republican and Independent PCs respectively. Qualitatively, the degree distributions have similar patterns for self-reported and estimated PCs with the same ideology. As a robustness check, we redo the analysis above, with each connection weighted by transfer amount. Specifically, for each PC, we calculate its total amount of transfer to and from (estimated) Democratic, Republican, and Independent PCs respectively, and then plot the distributions of these numbers for different groups of PCs. The histograms are shown in Figures 2.15, 2.16, and 2.17, respectively for Democratic, Republican and Independent PCs. Again, the distributions are similar for self-reported and estimated PCs with the same ideology. These results demonstrate that PCs without self-report, classified according to our estimation results, behave similarly to PCs with the corresponding self reports of affiliations.

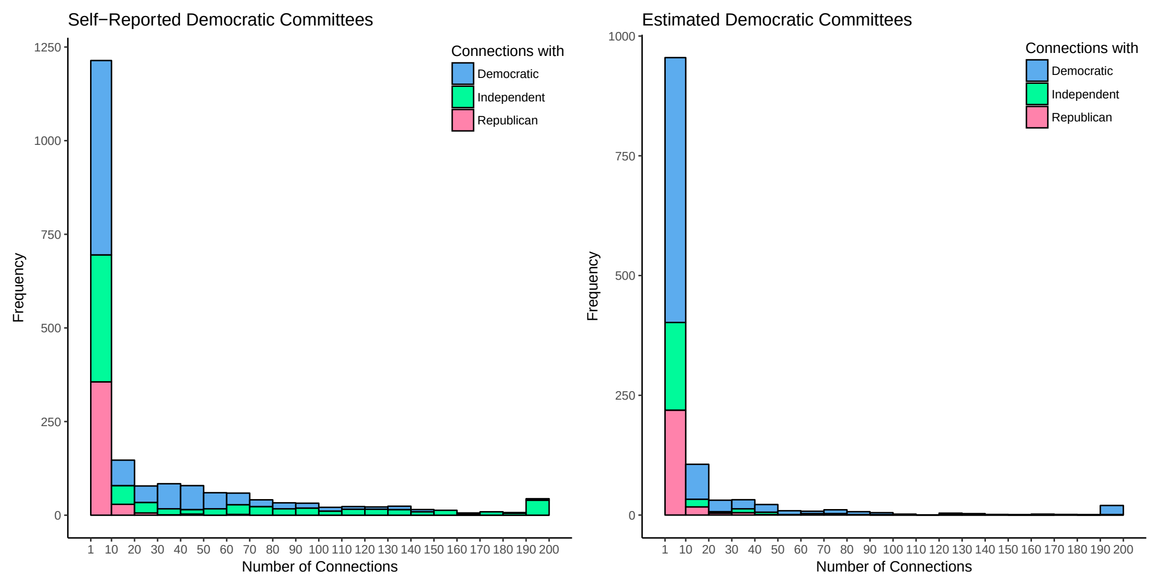

Figure 2.12: Self-reported vs. Estimated Democratic PCs: Distributions of Number of Connections

Note: These figures only include observations with positive number of connections. Observations with more than 200 connections are plotted at 200.

Estimation Results: Analyzing the Posterior Distribution of PCs' Political Ideologies

Now we analyze the posterior distribution of political ideology. Figure 2.18 plots, for PCs whose estimates are the same as the self reports (i.e., the reported ideology has the highest posterior probability), the distribution of the differences between the highest and the second highest posterior probability. These differences concentrate around 1, meaning the posterior probabilities concentrate on the self-reported affiliation. This

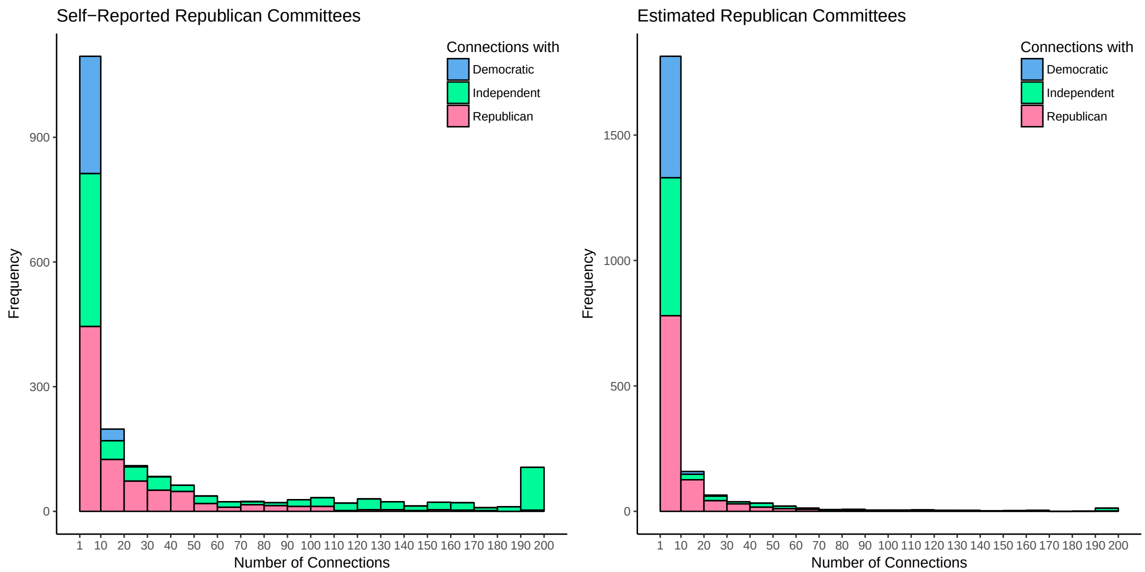

Figure 2.13: Self-reported vs. Estimated Republican PCs: Distributions of Number of Connections

Note: These figures only include observations with positive number of connections. Observations with more than 200 connections are plotted at 200.

confirms that we do not obtain these classifications by luck. We do a similar analysis in Figure 2.19, for PCs whose estimates differ from the self reports, by plotting the distributions of the differences between the highest posterior probability and the posterior probability of the reported ideology. These distributions have larger spread, but still a mass around 1. This indicates that these are not "near misses": our estimate strongly favors an ideology different from the PC's self-reported ideology.

Diagnostics of the Discrepancy. We further investigate the reason for the discrepancy between our estimates and the self reports by comparing PCs with different estimated ideology but the same self reported affiliation.

First of all, their financial conditions are different. Table 2.15 shows that self-reported Democratic PCs that are estimated as Republican are mostly PCs with high

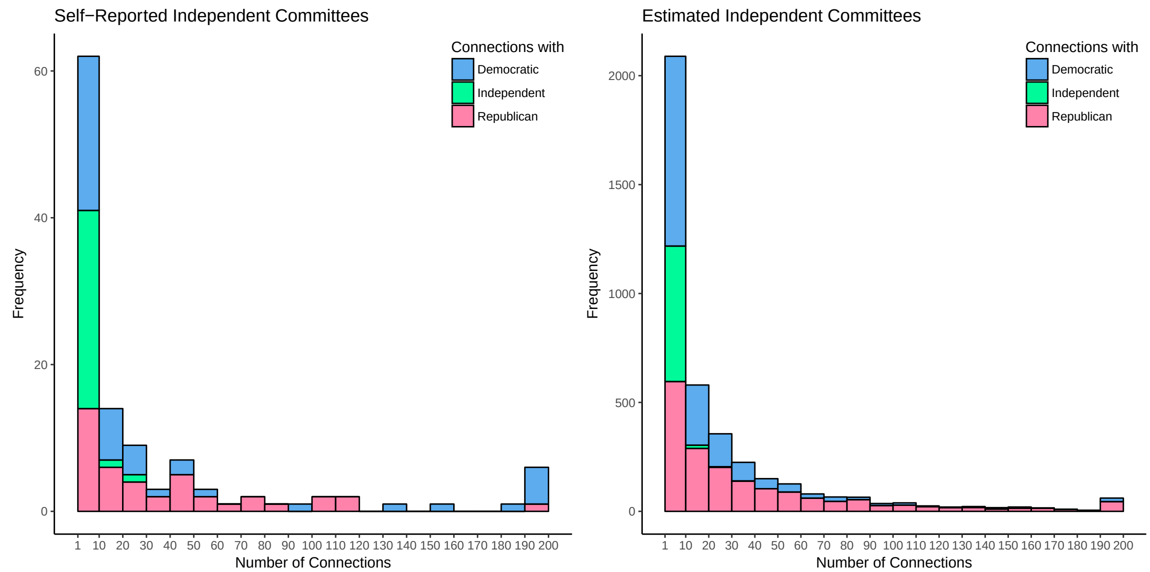

Figure 2.14: Self-reported vs. Estimated Independent PCs: Distributions of Number of Connections

Note: These figures only include observations with positive number of connections. Observations with more than 200 connections are plotted at 200.

budget, and self-reported Democratic PCs that are estimated as Independent are mostly PCs with low budget. Similarly, among self-reported Republican PCs, the ones estimated to be Democratic or Independent are mostly PCs with lower budget.

	Estimated Dem	Estimated Rep	Estimated Ind
Self-Reported Dem	32.26	206.52	3.50
Self-Reported Rep	10.00	77.96	6.77
Self-Reported Ind	31.00	3.00	46.00

Table 2.15: Median Budget (in \$1,000)

Second, their contribution patterns are different. For each PC i, we define DemShare_i as its share of connections with (estimated) Democratic PCs $\text{DemShare}_i = \frac{\text{numDem}_i}{\text{numDem}_i + \text{numRep}_i + \text{numInd}_i}$, and similarly for RepShare_i and IndShare_i. In Table 2.16, we compare the means of DemShare, RepShare, and IndShare for different groups of

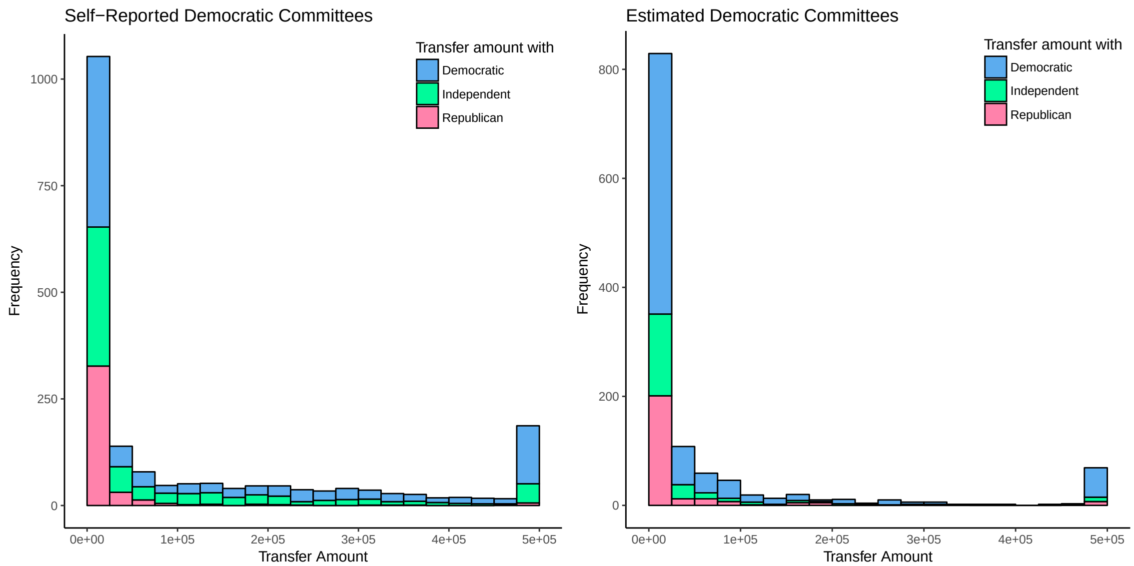

Figure 2.15: Self-reported vs. Estimated Democratic PCs: Distributions of Transfer Amount

Note: Bin size is 25,000. Observations with more than $500,000 transfer to and from one class of committees are plotted at $500,000.

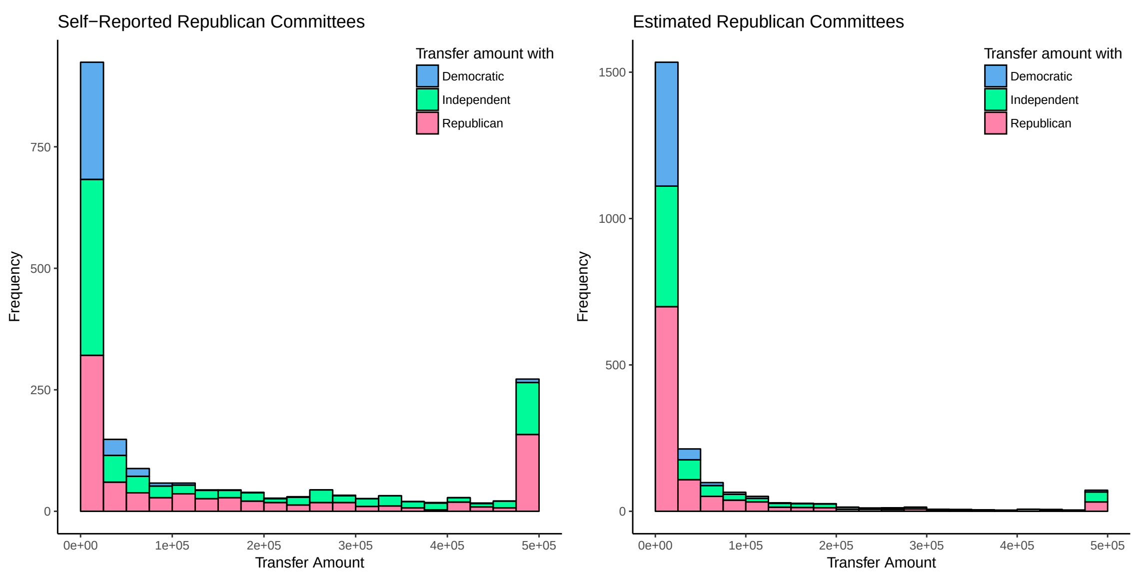

Figure 2.16: Self-reported vs. Estimated Republican PCs: Distributions of Transfer Amount

Note: Bin size is 25,000. Observations with more than $500,000 transfer to and from one class of committees are plotted at $500,000.

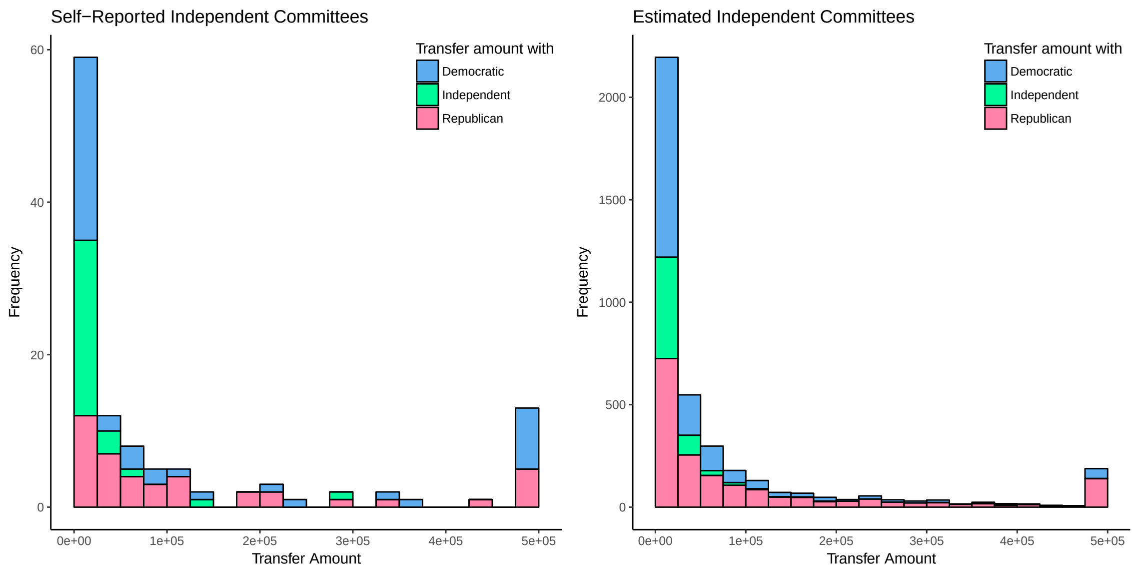

Figure 2.17: Self-reported vs. Estimated Independent PCs: Distributions of Transfer Amount

Note: Bin size is 25,000. Observations with more than \$500,000 transfer to and from one class of committees are plotted at \$500,000.

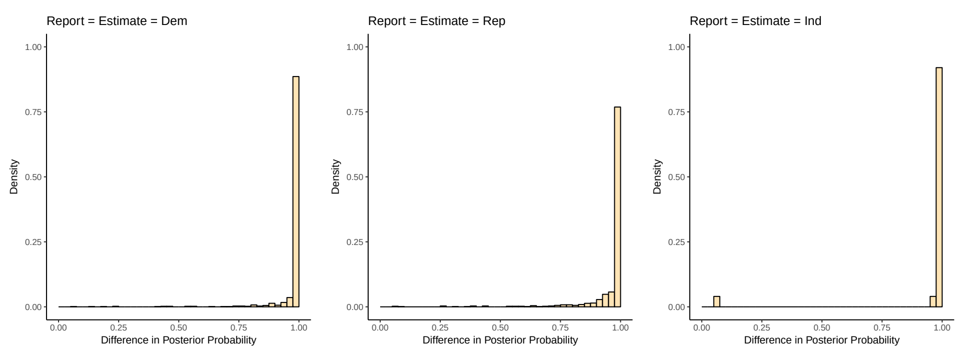

Figure 2.18: Distribution of Difference in the Posterior Probability of Ideology

Note: Horizontal axis is the difference between the highest posterior probability (i.e., the posterior probability of the self-reported ideology) and the second highest posterior probability. Bin size is 0.025.

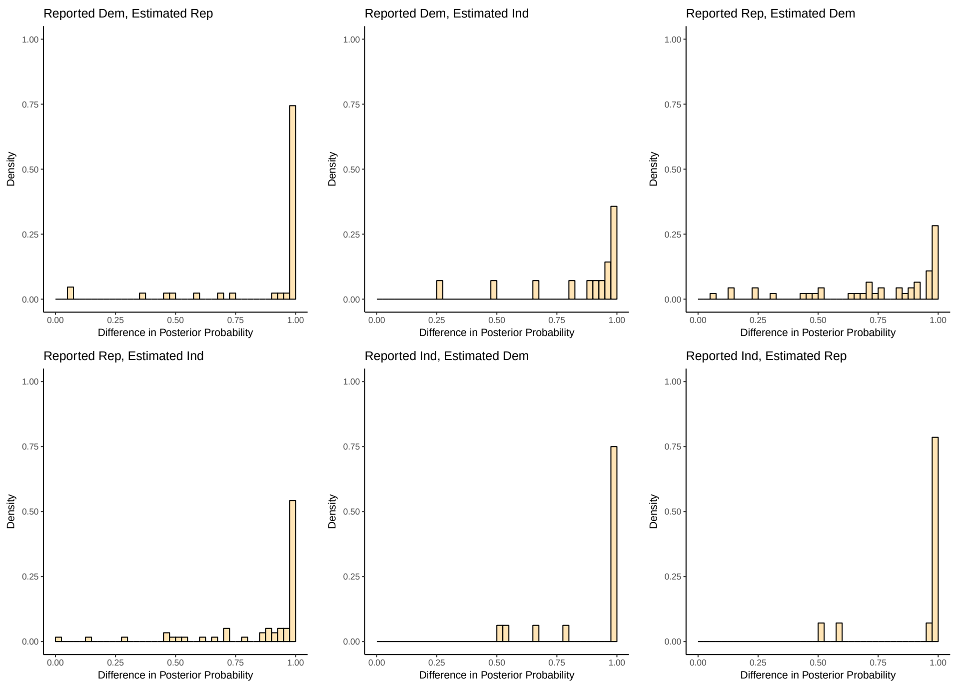

Figure 2.19: Distribution of Difference in the Posterior Probability of Ideology
Note: Horizontal axis is the difference between the highest posterior probability and the posterior probability of the self-reported ideology. Bin size is 0.025.

PCs. As a robustness check, we also calculate the shares in terms of transfer amount and present the results in the same table. The table supports our categorization of some PCs as Republican (Democratic) whose self-reports are Democratic (Republican) because they have significantly higher shares of connections with the Republican (Democratic) PCs, a pattern exhibited by all estimated Republican (Democratic) PCs.

Next, we will discuss the discrepancy in more detail case by case. The discrepancies will fall into one of six cases, as shown in the panel label of Figure 2.20. Figure 2.20 depicts the distribution of the number of connections with PCs with different ideologies according to our estimates.

	Number of Connection			Transfer Amount		
	Dem Share	Rep Share	Ind Share	Dem Share	Rep Share	Ind Share
Reported Dem, Estimated Rep	8.71%	19.65%	71.64%	28.36%	20.51%	51.13%
All PCs Estimated as Dem	72.58%	4.40%	23.02%	78.24%	4.15%	17.62%
Reported Rep, Estimated Dem	49.99%	13.44%	36.58%	54.56%	21.20%	24.24%
All PCs Estimated as Rep	6.69%	52.98%	40.33%	6.01%	55.77%	38.22%

Table 2.16: Mean of DemShare, RepShare, and IndShare

In the first case, some self-reported Democratic PCs are estimated to be Republican. In the data, these PCs have a small number of connections with PCs that self reported affiliations, most of which are Democratic. However, they are mostly connected with PCs without self reported affiliations, most of which are Independent by our estimates. As our estimate $\boldsymbol{\beta}$ suggests, Republican PCs have the highest connection probability with Independent PCs. Therefore, although they self reported to be Democratic and are connected with few self-reported Republican or Independent PCs, their overall contribution patterns are close to that of the Republican PCs. In the degree distribution presented in Figure 2.20 for this case, it can be seen that the tail distribution is very close to that of the self-reported Republican PCs in Figure 2.13.

In the second case, some self-reported Democratic PCs are estimated as Independent. Only one of these PCs has one connection with another Independent PC, and the rest have no connection with Independent PCs. According to estimate $\boldsymbol{\beta}$, classifying them as Independent, rather than Democratic, better rationalize this pattern because Independent-Independent connection probability is lower than Democratic-Independent connection probability.

In the third case, some self-reported Republican PCs are estimated as Democratic. These PCs have similar number of connections to all three classes of PCs, with slightly more Democratic connections. Their number of connections to the Republican and the Independent PCs are not jointly high enough to be estimated as Republican.

In the fourth case, some self-reported Republican PCs are estimated as Independent. These PCs have more connections with Republican than Democratic PCs, but not large enough connections with Independent PCs to be estimated as Republican. In other words, they do not exhibit the heavy tail on connection with Independent PCs which is observed for other self-reported Republican PCs.

In the fifth case, some self-reported Independent PCs are estimated as Democratic. These PCs' numbers of connections with Democratic PCs significantly outweigh that with the Republican and Independent PCs. There are not large enough connections with Republican PCs to be estimated as Independent.

In the sixth case, some self-reported Independent PCs are estimated as Republican. These PCs do not have large enough connections with Democratic PCs to be estimated as Independent.

Discussion. Here we present and analyze the discrepancy between our estimate and the self-report, in an attempt to better understand the implications of our model and algorithm. We are not making a claim that these PCs strategically misreported their party affiliations. We show that under our model, their contribution patterns to other PCs are different from the majority of the PCs self reporting the same affiliation as they did. Simplifications in our model are also potential reasons for the discrepancy. For example, our model does not capture all aspects of the incentives in political contribution between PCs. Additionally, we only study the contributions among PCs, and do leave out other campaign activities such as collection of individual contribution and independent expenditures. In Appendix 4.7 we list the PCs that self reported to be Democratic (Republican), but are estimated to be Republican (Democratic).

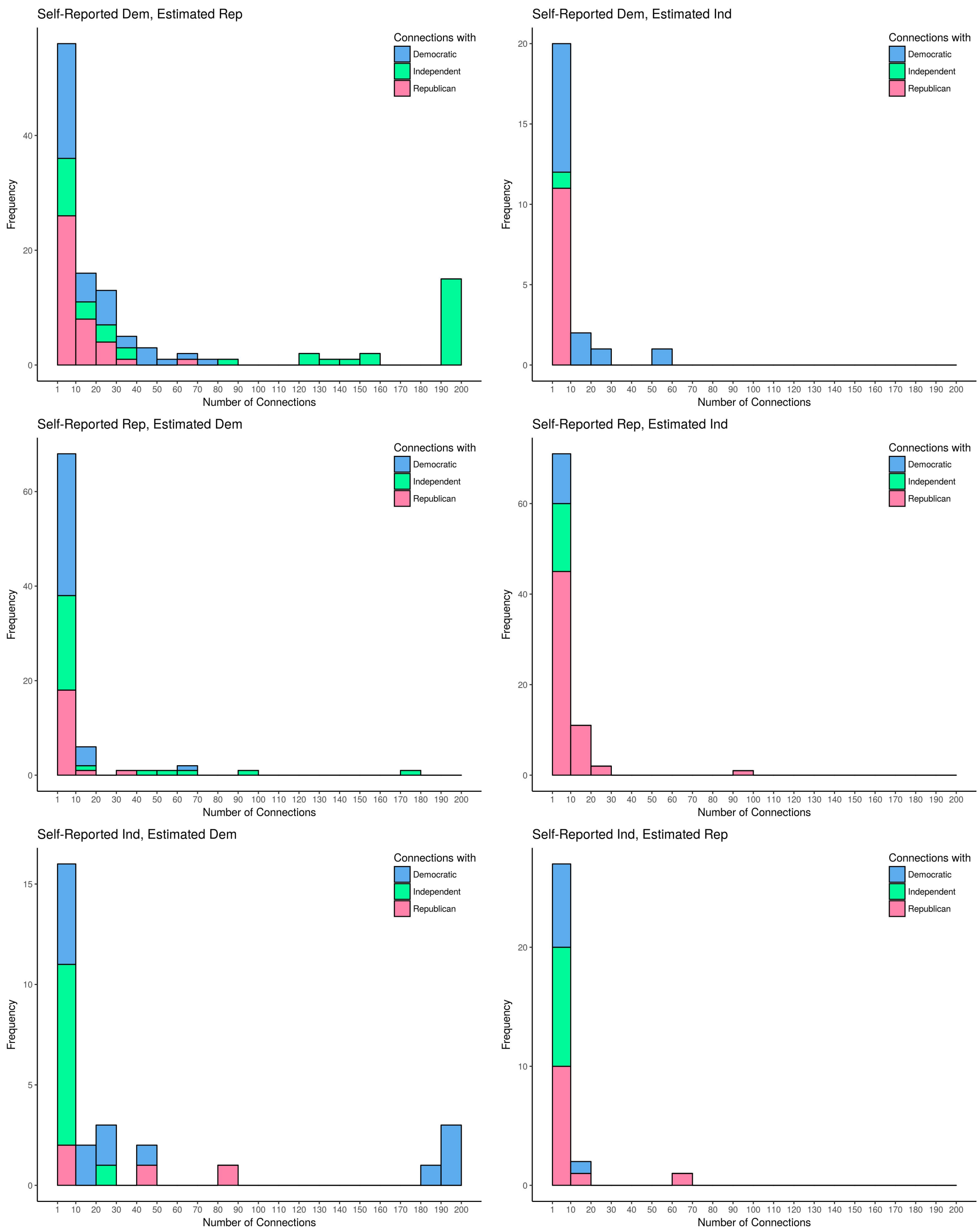

Figure 2.20: Distributions of Number of Connections with Different Committees
Note: Observations with more than 200 connections are plotted at 200.

2.7 Conclusion

About two thirds of the political committees registered with the Federal Election Commission do not self identify their party affiliations. In this paper we propose and implement a novel Bayesian approach to infer about the ideological affiliations of political committees based on the network of the financial contributions among them. In Monte Carlo simulations, we demonstrate that our estimation algorithm achieves very high accuracy in recovering their latent ideological affiliations when the pairwise difference in ideology groups' connection patterns satisfy a condition known as the Chernoff-Hellinger divergence criterion. We illustrate our approach using the campaign finance record in 2003-2004 election cycle. Using the posterior mode to categorize the ideological affiliations of the political committees, our estimates match the self reported ideology for 94.36% of those committees who self reported to be Democratic and 89.49% of those committees who self reported to be Republican.

Since PCs are required to report to the FEC the financial contributions among each other, our proposed methods to infer the ideological affiliations of political committees via financial contributions network can be implemented readily. To the extent that the estimated ideologies for the PCs are close to their true latent ideologies, our estimated PC ideology can fill the missing "ideologies" problem for researchers who are interested in studying *individuals'* political contribution patterns using FEC's "Contributions by Individuals" data. Moreover, since our estimation methods can be implemented using data from only one election cycle, we can estimate the ideological affiliations of the same PCs using data from different election cycles. This would allow us to study the possible evolutions of ideological affiliations of PCs over time. We can also exploit the permanent presence of the national committees such as Democratic

National Committee (DNC) and Republican National Committee (RNC) and use their network links as a vehicle to study the possible changes of party platforms that are not necessarily reflected in official documents. These are exciting areas for future research.

Chapter 3

Appendix for Chapter 1

3.1 Bargained Wage

Unemployed worker's value function is

$$\begin{aligned}\rho V_0(a,\Gamma) =& U(ab) \\ &+ \lambda_0 \int_{p_0(a,\Gamma)}^{p^{max}} [V_1(a,\phi_0(a,x,\Gamma),x,\Gamma) - V_0(a,\Gamma)]dF(x) \\ &+ \delta\pi_0 \int_{p_0(a,\Gamma)}^{p^{max}} [V_1(a,\phi_0(a,y,\Gamma),y,\Gamma) - V_0(a,\Gamma)]d\Gamma_0(y) \\ &+ \lambda_1\pi_1 \int_{p_0(a,\Gamma)}^{p^{max}} \int_{p_0(a,\Gamma)}^{y} [V_1(a,\phi_0(a,x,\Gamma),x,\Gamma) - V_0(a,\Gamma)]dF(x)d\Gamma_0(y) \\ &+ \lambda_1\pi_1 \int_{p_0(a,\Gamma)}^{p^{max}} \int_{y}^{p^{max}} [V_1(a,\phi_0(a,y,\Gamma),y,\Gamma) - V_0(a,\Gamma)]dF(x)d\Gamma_0(y),\end{aligned} \tag{3.1}$$

where state variable $\Gamma = \{\Gamma_0, \Gamma_1\}$, and $p_0(a,\Gamma)$ is the reservation productivity, i.e.

$$V_0(a,\Gamma) = V_1(a, ap_0(a,\Gamma), p_0(a,\Gamma), \Gamma). \tag{3.2}$$

Simplify the expression using the property of ϕ_0 implied from the bargaining process,

$$\begin{aligned}
\rho V_0(a,\Gamma) =& U(ab) \\
&+ \lambda_0 \int_{p_0(a,\Gamma)}^{p^{max}} \beta[V_1(a,ax,x,\Gamma) - V_0(a,\Gamma)]dF(x) \\
&+ \delta\pi_0 \int_{p_0(a,\Gamma)}^{p^{max}} \beta[V_1(a,ay,y,\Gamma) - V_0(a,\Gamma)]d\Gamma_0(y) \\
&+ \lambda_1\pi_1 \int_{p_0(a,\Gamma)}^{p^{max}} \int_{p_0(a,\Gamma)}^{y} \beta[V_1(a,ax,x,\Gamma) - V_0(a,\Gamma)]dF(x)d\Gamma_0(y) \\
&+ \lambda_1\pi_1 \int_{p_0(a,\Gamma)}^{p^{max}} \int_{y}^{p^{max}} \beta[V_1(a,ay,y,\Gamma) - V_0(a,\Gamma)]dF(x)d\Gamma_0(y).
\end{aligned} \tag{3.3}$$

Employed worker's value function is

$$
\begin{aligned}
&\rho V_1(a,w,p,\Gamma)\\
&= U(w) + \delta[V_0(a,\Gamma) - V_1(a,w,p,\Gamma)]\\
&+ \lambda_1 \int_{q(a,w,p,\Gamma)}^{p} [V_1(a,\phi_1(a,x,p,\Gamma),p,\Gamma) - V_1(a,w,p,\Gamma)]dF(x)\\
&+ \lambda_1 \int_{p}^{p^{max}} [V_1(a,\phi_1(a,p,x,\Gamma),x,\Gamma) - V_1(a,w,p,\Gamma)]dF(x)\\
&+ \delta\pi_0 \int_{q(a,w,p,\Gamma)}^{p} [V_1(a,\phi_1(a,y,p,\Gamma),p,\Gamma) - V_1(a,w,p,\Gamma)]d\Gamma_1(y)\\
&+ \delta\pi_0 \int_{p}^{p^{max}} [V_1(a,\phi_1(a,p,y,\Gamma),y,\Gamma) - V_1(a,w,p,\Gamma)]d\Gamma_1(y)\\
&+ \lambda_1\pi_1 \int_{q(a,w,p,\Gamma)}^{p} \int_{q(a,w,p,\Gamma)}^{y} [V_1(a,\phi_1(a,x,p,\Gamma),p,\Gamma) - V_1(a,w,p,\Gamma)]dF(x)d\Gamma_1(y)\\
&+ \lambda_1\pi_1 \int_{q(a,w,p,\Gamma)}^{p} \int_{y}^{p^{max}} [V_1(a,\phi_1(a,y,p,\Gamma),p,\Gamma) - V_1(a,w,p,\Gamma)]dF(x)d\Gamma_1(y)\\
&+ \lambda_1\pi_1 \int_{p}^{p^{max}} \int_{q(a,w,p,\Gamma)}^{p} [V_1(a,\phi_1(a,x,p,\Gamma),p,\Gamma) - V_1(a,w,p,\Gamma)]dF(x)d\Gamma_1(y)\\
&+ \lambda_1\pi_1 \int_{p}^{p^{max}} \int_{p}^{y} [V_1(a,\phi_1(a,p,x,\Gamma),x,\Gamma) - V_1(a,w,p,\Gamma)]dF(x)d\Gamma_1(y)\\
&+ \lambda_1\pi_1 \int_{p}^{p^{max}} \int_{y}^{p^{max}} [V_1(a,\phi_1(a,p,y,\Gamma),y,\Gamma) - V_1(a,w,p,\Gamma)]dF(x)d\Gamma_1(y),
\end{aligned}
\tag{3.4}
$$

where $q(a,w,p,\Gamma)$ is the cutoff productivity for wage raise, i.e.,

$$\phi_1(a,q(a,w,p,\Gamma),p,\Gamma) = w, \tag{3.5}$$

or equivalently,

$$V_1(a,w,p,\Gamma) = (1-\beta)V_1(a,aq(a,w,p,\Gamma),q(a,w,p,\Gamma),\Gamma) + \beta V_1(a,ap,p,\Gamma). \tag{3.6}$$

Simplify the expression using the property of ϕ_1 implied from the bargaining process,

$$
\begin{aligned}
&\rho V_1(a,w,p,\Gamma)\\
&= U(w) + \delta[V_0(a,\Gamma) - V_1(a,w,p,\Gamma)]\\
&+ \lambda_1 \int_{q(a,w,p,\Gamma)}^{p} \Big\{[\beta V_1(a,ap,p,\Gamma) + (1-\beta)V_1(a,ax,x,\Gamma)] - V_1(a,w,p,\Gamma)\Big\} dF(x)\\
&+ \lambda_1 \int_{p}^{p^{max}} \Big\{[\beta V_1(a,ax,x,\Gamma) + (1-\beta)V_1(a,ap,p,\Gamma)] - V_1(a,w,p,\Gamma)\Big\} dF(x)\\
&+ \delta\pi_0 \int_{q(a,w,p,\Gamma)}^{p} \Big\{[\beta V_1(a,ap,p,\Gamma) + (1-\beta)V_1(a,ay,y,\Gamma)] - V_1(a,w,p,\Gamma)\Big\} d\Gamma_1(y)\\
&+ \delta\pi_0 \int_{p}^{p^{max}} \Big\{[\beta V_1(a,ay,y,\Gamma) + (1-\beta)V_1(a,ap,p,\Gamma)] - V_1(a,w,p,\Gamma)\Big\} d\Gamma_1(y)\\
&+ \lambda_1\pi_1 \int_{q(a,w,p,\Gamma)}^{p} \int_{q(a,w,p,\Gamma)}^{y} \Big\{[\beta V_1(a,ap,p,\Gamma) + (1-\beta)V_1(a,ax,x,\Gamma)] - V_1(a,w,p,\Gamma)\Big\} dF(x)d\Gamma_1(y)\\
&+ \lambda_1\pi_1 \int_{q(a,w,p,\Gamma)}^{p} \int_{y}^{p^{max}} \Big\{[\beta V_1(a,ap,p,\Gamma) + (1-\beta)V_1(a,ay,y,\Gamma)] - V_1(a,w,p,\Gamma)\Big\} dF(x)d\Gamma_1(y)\\
&+ \lambda_1\pi_1 \int_{p}^{p^{max}} \int_{q(a,w,p,\Gamma)}^{p} \Big\{[\beta V_1(a,ap,p,\Gamma) + (1-\beta)V_1(a,ax,x,\Gamma)] - V_1(a,w,p,\Gamma)\Big\} dF(x)d\Gamma_1(y)\\
&+ \lambda_1\pi_1 \int_{p}^{p^{max}} \int_{p}^{y} \Big\{[\beta V_1(a,ax,x,\Gamma) + (1-\beta)V_1(a,ap,p,\Gamma)] - V_1(a,w,p,\Gamma)\Big\} dF(x)d\Gamma_1(y)\\
&+ \lambda_1\pi_1 \int_{p}^{p^{max}} \int_{y}^{p^{max}} \Big\{[\beta V_1(a,ay,y,\Gamma) + (1-\beta)V_1(a,ap,p,\Gamma)] - V_1(a,w,p,\Gamma)\Big\} dF(x)d\Gamma_1(y).
\end{aligned}
\tag{3.7}
$$

First, define $\bar{F}(\cdot) = 1 - F(\cdot), \bar{\Gamma}_0(\cdot) = \Gamma_0(p^{\max}) - \Gamma_0(\cdot), \bar{\Gamma}_1(\cdot) = \Gamma_1(p^{\max}) - \Gamma_1(\cdot)$.

Simplify (3.7) by integrating by part, using property (3.6) and collecting terms of $V_1(a,w,p,\Gamma)$,

$$
\begin{aligned}
(\rho+\delta)V_1(a,w,p,\Gamma) =& U(w) + \delta V_0(a,\Gamma) \\
&+ \lambda_1(1-\beta)\int_{q(a,w,p,\Gamma)}^{p} \bar{F}(x)\frac{dV_1(a,ax,x,\Gamma)}{dx}dx \\
&+ \lambda_1\beta\int_{p}^{p^{max}} \bar{F}(x)\frac{dV_1(a,ax,x,\Gamma)}{dx}dx \\
&+ \delta\pi_0(1-\beta)\int_{q(a,w,p,\Gamma)}^{p} \bar{\Gamma}_1(y)\frac{dV_1(a,ay,y,\Gamma)}{dy}dy \\
&+ \delta\pi_0\beta\int_{p}^{p^{max}} \bar{\Gamma}_1(y)\frac{dV_1(a,ay,y,\Gamma)}{dy}dy \\
&+ \lambda_1\pi_1(1-\beta)\int_{q(a,w,p,\Gamma)}^{p}\int_{q(a,w,p,\Gamma)}^{y} \bar{F}(x)\frac{dV_1(a,ax,x,\Gamma)}{dx}dx d\Gamma_1(y) \\
&+ \lambda_1\pi_1(1-\beta)\int_{p}^{p^{max}}\int_{q(a,w,p,\Gamma)}^{p} \bar{F}(x)\frac{dV_1(a,ax,x,\Gamma)}{dx}dx d\Gamma_1(y) \\
&+ \lambda_1\pi_1\beta\int_{p}^{p^{max}}\int_{p}^{y} \bar{F}(x)\frac{dV_1(a,ax,x,\Gamma)}{dx}dx d\Gamma_1(y).
\end{aligned}
\tag{3.8}
$$

For the special case where $w = ap$, $q(a, ap, p, \Gamma) = p$, and (3.8) becomes

$$
\begin{aligned}
(\rho+\delta)V_1(a,ap,p,\Gamma) =& U(ap) + \delta V_0(a,\Gamma) \\
&+ \lambda_1\beta\int_{p}^{p^{max}} \bar{F}(x)\frac{dV_1(a,ax,x,\Gamma)}{dx}dx \\
&+ \delta\pi_0\beta\int_{p}^{p^{max}} \bar{\Gamma}_1(y)\frac{dV_1(a,ay,y,\Gamma)}{dy}dy \\
&+ \lambda_1\pi_1\beta\int_{p}^{p^{max}}\int_{p}^{y} \bar{F}(x)\frac{dV_1(a,ax,x,\Gamma)}{dx}dx d\Gamma_1(y).
\end{aligned}
\tag{3.9}
$$

Differentiate (3.9) with respect to p,

$$
\frac{dV_1(a,ap,p,\Gamma)}{dp} = \frac{aU'(ap)}{\rho+\delta+\lambda_1\beta\bar{F}(p)+[\lambda_1\pi_1\beta\bar{F}(p)+\delta\pi_0\beta]\bar{\Gamma}_1(p)}. \tag{3.10}
$$

To solve for bargained wage, note that $\phi_1(a, p^L, p^H, \Gamma)$ is equal to ϕ_1^* that solves

$$
V_1(a,\phi_1^*,p^H,\Gamma) = \beta V_1(a,ap^H,p^H,\Gamma) + (1-\beta)V_1(a,ap^L,p^L,\Gamma). \tag{3.11}
$$

Plug in expressions (3.8) and (3.9) and use the property that $q(a, \phi_1^*, p^H, \Gamma) = p^L$,

$$\begin{aligned} U(\phi_1(a, p^L, p^H, \Gamma)) =& \beta U(ap^H) + (1-\beta)U(ap^L) \\ & - \delta\pi_0(1-\beta)^2 \int_{p^L}^{p^H} \bar{\Gamma}_1(y) \frac{dV_1(a, ay, y, \Gamma)}{dy} dy \\ & - \lambda_1(1-\beta)^2 \Bigg\{ \int_{p^L}^{p^H} \bar{F}(x) \frac{dV_1(a, ax, x, \Gamma)}{dx} dx \\ & + \pi_1 \int_{p^H}^{p^{max}} \int_{p^L}^{p^H} \bar{F}(x) \frac{dV_1(a, ax, x, \Gamma)}{dx} dx d\Gamma_1(y) \\ & + \pi_1 \int_{p^L}^{p^H} \int_{p^L}^{y} \bar{F}(x) \frac{dV_1(a, ax, x, \Gamma)}{dx} dx d\Gamma_1(y) \Bigg\}, \end{aligned} \tag{3.12}$$

where $\frac{dV_1(a,ax,x,\Gamma)}{dx}$ is given in (3.10).

Simplify (3.3) by integrating by part,

$$\begin{aligned} \rho V_0(a, \Gamma) =& U(ab) \\ & + \lambda_0 \beta \int_{p_0(a,\Gamma)}^{p^{max}} \bar{F}(x) \frac{dV_1(a, ax, x, \Gamma)}{dx} dx + \delta\pi_0\beta \int_{p_0(a,\Gamma)}^{p^{max}} \bar{\Gamma}_0(y) \frac{dV_1(a, ay, y, \Gamma)}{dy} dy \\ & + \lambda_1\pi_1\beta \int_{p_0(a,\Gamma)}^{p^{max}} \int_{p_0(a,\Gamma)}^{y} \bar{F}(x) \frac{dV_1(a, ax, x, \Gamma)}{dx} dx d\Gamma_0(y). \end{aligned} \tag{3.13}$$

To solve for unemployed worker's reservation productivity, note that $p_0(a, \Gamma)$ is equal to p_0^* that solves

$$V_1(a, ap_0^*, p_0^*, \Gamma) = V_0(a, \Gamma). \tag{3.14}$$

Plugging in expressions (3.13) and (3.9) gives an implicit function for $p_0(a, \Gamma)$

$$\begin{aligned} U(ap_0(a, \Gamma)) =& U(ab) + \beta(\lambda_0 - \lambda_1) \int_{p_0(a,\Gamma)}^{p^{max}} \bar{F}(x) \frac{dV_1(a, ax, x, \Gamma)}{dx} dx \\ & + \delta\pi_0\beta \int_{p_0(a,\Gamma)}^{p^{max}} (\bar{\Gamma}_0(y) - \bar{\Gamma}_1(y)) \frac{dV_1(a, ay, y, \Gamma)}{dy} dy \\ & + \lambda_1\pi_1\beta \int_{p_0(a,\Gamma)}^{p^{max}} (\bar{\Gamma}_0(x) - \bar{\Gamma}_1(x)) \bar{F}(x) \frac{dV_1(a, ay, y, \Gamma)}{dx} dx, \end{aligned} \tag{3.15}$$

where $\frac{dV_1(a,ax,x,\Gamma)}{dx}$ is given in (3.10).

By the definition of $p_0(a,\Gamma)$,

$$\phi_0(a,p,\Gamma) = \phi_1(a,p_0(a,\Gamma),p,\Gamma). \tag{3.16}$$

In summary, the bargained wage is characterized by (3.10), (3.12), (3.15), and (3.16).

For CRRA utility with $\alpha \neq 1$, (3.12) simplifies to

$$\begin{aligned}
\ln\phi_1(a,p^L,p^H,\Gamma) = &\ln a + \frac{1}{1-\alpha}\ln\Bigg\{\beta(P^H)^{1-\alpha} + (1-\beta)(P^L)^{1-\alpha} \\
&- (1-\alpha)\delta\pi_0(1-\beta)^2\int_{p^L}^{p^H}\frac{\bar{\Gamma}_1(y)y^{-\alpha}}{\rho+\delta+\lambda_1\beta\bar{F}(y)+[\lambda_1\pi_1\beta\bar{F}(y)+\delta\pi_0\beta]\bar{\Gamma}_1(y)}dy \\
&- (1-\alpha)\lambda_1(1-\beta)^2\Big[\int_{p^L}^{p^H}\frac{\bar{F}(x)x^{-\alpha}}{\rho+\delta+\lambda_1\beta\bar{F}(x)+[\lambda_1\pi_1\beta\bar{F}(x)+\delta\pi_0\beta]\bar{\Gamma}_1(x)}dx \\
&+ \pi_1\int_{p^H}^{p^{max}}\int_{p^L}^{p^H}\frac{\bar{F}(x)x^{-\alpha}}{\rho+\delta+\lambda_1\beta\bar{F}(x)+[\lambda_1\pi_1\beta\bar{F}(x)+\delta\pi_0\beta]\bar{\Gamma}_1(x)}dxd\Gamma_1(y) \\
&+ \pi_1\int_{p^L}^{p^H}\int_{p^L}^{y}\frac{\bar{F}(x)x^{-\alpha}}{\rho+\delta+\lambda_1\beta\bar{F}(x)+[\lambda_1\pi_1\beta\bar{F}(x)+\delta\pi_0\beta]\bar{\Gamma}_1(x)}dxd\Gamma_1(y)\Big]\Bigg\},
\end{aligned} \tag{3.17}$$

and (3.15) simplifies to

$$\begin{aligned}
\ln p_0(a,\Gamma) = &\frac{1}{1-\alpha}\ln\Bigg\{b^{1-\alpha} + (1-\alpha)\beta(\lambda_0-\lambda_1)\int_{p_0(a,\Gamma)}^{p^{max}}\frac{\bar{F}(x)x^{-\alpha}}{\rho+\delta+\lambda_1\beta\bar{F}(x)+[\lambda_1\pi_1\beta\bar{F}(x)+\delta\pi_0\beta]\bar{\Gamma}_1(x)}dx \\
&+ (1-\alpha)\delta\pi_0\beta\int_{p_0(a,\Gamma)}^{p^{max}}\frac{(\bar{\Gamma}_0(y)-\bar{\Gamma}_1(y))y^{-\alpha}}{\rho+\delta+\lambda_1\beta\bar{F}(y)+[\lambda_1\pi_1\beta\bar{F}(y)+\delta\pi_0\beta]\bar{\Gamma}_1(y)}dy \\
&+ (1-\alpha)\lambda_1\pi_1\beta\int_{p_0(a,\Gamma)}^{p^{max}}\frac{(\bar{\Gamma}_0(x)-\bar{\Gamma}_1(x))\bar{F}(x)x^{-\alpha}}{\rho+\delta+\lambda_1\beta\bar{F}(x)+[\lambda_1\pi_1\beta\bar{F}(x)+\delta\pi_0\beta]\bar{\Gamma}_1(x)}dx\Bigg\};
\end{aligned} \tag{3.18}$$

for CRRA utility with $\alpha = 1$, (3.12) simplifies to

$$\begin{aligned}
\ln \phi_1(a, p^L, p^H, \Gamma) = {} & \ln a + \beta \ln p^H + (1-\beta)\ln p^L \\
& - \delta\pi_0(1-\beta)^2 \int_{p^L}^{p^H} \frac{\bar{\Gamma}_1(y)y^{-1}}{\rho+\delta+\lambda_1\beta\bar{F}(y)+[\lambda_1\pi_1\beta\bar{F}(y)+\delta\pi_0\beta]\bar{\Gamma}_1(y)}dy \\
& - \lambda_1(1-\beta)^2\Big[\int_{p^L}^{p^H} \frac{\bar{F}(x)x^{-1}}{\rho+\delta+\lambda_1\beta\bar{F}(x)+[\lambda_1\pi_1\beta\bar{F}(x)+\delta\pi_0\beta]\bar{\Gamma}_1(x)}dx \\
& + \pi_1 \int_{p^H}^{p^{max}} \int_{p^L}^{p^H} \frac{\bar{F}(x)x^{-1}}{\rho+\delta+\lambda_1\beta\bar{F}(x)+[\lambda_1\pi_1\beta\bar{F}(x)+\delta\pi_0\beta]\bar{\Gamma}_1(x)}dxd\Gamma_1(y) \\
& + \pi_1 \int_{p^L}^{p^H} \int_{p^L}^{y} \frac{\bar{F}(x)x^{-1}}{\rho+\delta+\lambda_1\beta\bar{F}(x)+[\lambda_1\pi_1\beta\bar{F}(x)+\delta\pi_0\beta]\bar{\Gamma}_1(x)}dxd\Gamma_1(y)\Big],
\end{aligned} \tag{3.19}$$

and (3.15) simplifies to

$$\begin{aligned}
\ln p_0(a,\Gamma) = {} & \ln b + \beta(\lambda_0-\lambda_1)\int_{p_0(a,\Gamma)}^{p^{max}} \frac{\bar{F}(x)x^{-1}}{\rho+\delta+\lambda_1\beta\bar{F}(x)+[\lambda_1\pi_1\beta\bar{F}(x)+\delta\pi_0\beta]\bar{\Gamma}_1(x)}dx \\
& + \delta\pi_0\beta \int_{p_0(a,\Gamma)}^{p^{max}} \frac{(\bar{\Gamma}_0(y)-\bar{\Gamma}_1(y))y^{-1}}{\rho+\delta+\lambda_1\beta\bar{F}(y)+[\lambda_1\pi_1\beta\bar{F}(y)+\delta\pi_0\beta]\bar{\Gamma}_1(y)}dy \\
& + \lambda_1\pi_1\beta \int_{p_0(a,\Gamma)}^{p^{max}} \frac{(\bar{\Gamma}_0(x)-\bar{\Gamma}_1(x))\bar{F}(x)x^{-1}}{\rho+\delta+\lambda_1\beta\bar{F}(x)+[\lambda_1\pi_1\beta\bar{F}(x)+\delta\pi_0\beta]\bar{\Gamma}_1(x)}dx.
\end{aligned} \tag{3.20}$$

3.2 Additional Table for Data Description

	Career Advancement		UE Transition		EE Transition		EU Transition	
#(Transition)	#(Individual)	%	#(Individual)	%	#(Individual)	%	#(Individual)	%
0	1,772	42.27	1,836	43.80	4,066	96.99	1,941	46.30
1	2,178	51.96	2171	51.79	125	2.98	2,085	49.73
2	231	5.51	179	4.27	1	0.02	161	3.84
3	10	0.24	6	0.14	0	0	5	0.12
4	1	0.02	0	0	0	0	0	0
Total	4,192	100.00	4,192	100.00	4,192	100.00	4,192	100.00

Table 3.1: Tabulation of Individuals' Numbers of Transitions

Notes: (1). A UE transition is defined as the transition from non-executive to executive. (2). An EE transition is defined as the transition from one executive job to another. (3). An EU transition is defined as the transition from executive to non-executive. (4). A career advancement is defined as either a UE or EE transition.

	(1)	(2)	(3)	(4)
Dependent variable: $\mathbb{1}$*(Career Advancement)*$\in \{0,1\}$				
$\mathbb{1}$ (Job Transition Among Executive Friends)	0.0470***			
	(0.0055)			
... from School		0.0166***		
		(0.0049)		
... from Work			0.0532***	
			(0.0050)	
... from Social Activity				0.0174**
				(0.0059)
Observations	33,536	33,536	33,536	33,536
Individual FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes

Table 3.2: Co-movement in Socially Connected Executives' Job Transitions

Notes: (1). This table summarizes the results of an OLS regression with year and individual fixed effects, where the dependent variable is whether an individual experiences a career advancement. (2). A career advancement is defined as either a transition from non-executive job to executive job (UE) or a transition between executive jobs with productivity increase (EE). (3). Variable $\mathbb{1}$ (Job Transition Among Executive Friends) is a dummy variable that equals one if any friend experiences a transition between executive jobs with productivity increase (EE) or a transition from executive job to non-executive job (EU) in the same year or the previous year. (4). The scope of friends varies for different network specifications. Column (1) uses the simplified network, which does not distinguish between types of social connections; column (2) uses the school network; column (3) uses the work network; and column (4) uses the social-activity network. (5). Other covariates include age^2, age^3, whether an individual was an executive in the previous year, and the productivity of the job in the previous year if the individual was an executive. (6). Standard errors are reported in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

3.3 Estimation Strategy

Likelihood L_1

The likelihood contribution of a worker i with m career advancements at time $t_1, t_2, ...t_m$ is

$$f_i(\boldsymbol{w_i};\boldsymbol{\theta}) = \frac{\exp\Big(-\frac{1}{2}(\ln \boldsymbol{w_i} - \ln \boldsymbol{\phi_i} - \boldsymbol{\mu_a})^T \boldsymbol{\Sigma_i}^{-1}(\ln \boldsymbol{w_i} - \ln \boldsymbol{\phi_i} - \boldsymbol{\mu_a})\Big)}{\sqrt{(2\pi)^m|\boldsymbol{\Sigma_i}|}}, \tag{3.21}$$

where

$$\ln \boldsymbol{w_i} = \begin{bmatrix} \ln w_{it_1} \\ ... \\ \ln w_{it_m} \end{bmatrix}_{m\times 1}, \quad \ln \boldsymbol{\phi_i} = \begin{bmatrix} \phi_1(1, p^L_{it_1}, p^H_{it_1}, \Gamma_{it_1}) \\ ... \\ \phi_1(1, p^L_{it_m}, p^H_{it_m}, \Gamma_{it_m}) \end{bmatrix}_{m\times 1}, \quad \boldsymbol{\mu_a} = \begin{bmatrix} \mu_a \\ ... \\ \mu_a \end{bmatrix}_{m\times 1}, \tag{3.22}$$

and

$$\boldsymbol{\Sigma_i} = \begin{bmatrix} \sigma_a^2 + \sigma_\epsilon^2 & ... & \sigma_a^2 \\ ... & ... & ... \\ \sigma_a^2 & ... & \sigma_a^2 + \sigma_\epsilon^2 \end{bmatrix}_{m\times m}. \tag{3.23}$$

For all the UE transitions, the model implies an additional constraint on the parameter space: $p^H_{it} \geq p_{0,it}(\boldsymbol{\theta})$, the observed productivity of the accepted job is no less than the calculated reservation productivity. I incorporate this constraint as a penalty term in the objective function. Instead of directly solving the constrained optimization problem

$$\begin{aligned} &\max \sum_i \ln f_i(\boldsymbol{w_i};\boldsymbol{\theta}) \\ &\text{s.t. } p^H_{it} \geq p_{0,it}(\boldsymbol{\theta}) \text{ for UE transition,} \end{aligned} \tag{3.24}$$

set the following objective function to solve an unconstrained optimization problem

$$\max \sum_i \ln f_i(\boldsymbol{w_i}; \boldsymbol{\theta}) - \kappa \sum_i g_i(\boldsymbol{\theta}), \tag{3.25}$$

where κ is a small positive number, and $g_i(\boldsymbol{\theta}) = \sum_{t \in \{t_1,\ldots,t_m\}} g_{it}(\boldsymbol{\theta})$ is a penalty function defined as

$$g_{it}(\boldsymbol{\theta}) = \begin{cases} 0 & \text{for EE transition,} \\ -\ln(p_{it}^H - p_{0,it}(\boldsymbol{\theta})) & \text{for UE transition where } p_{it}^H > p_{0,it}(\boldsymbol{\theta}), \\ +\infty & \text{for UE transition where } p_{it}^H \leq p_{0,it}(\boldsymbol{\theta}). \end{cases} \tag{3.26}$$

Function g is an interior penalty function, which guarantees that the solution to (3.25) satisfies the constraint. When κ is small enough, the solution to (3.25) will be close to the solution to original constrained problem (3.24).[70]

Likelihood L_2

The likelihood contribution of the productivity of a socially isolated worker i's first job p_i is

$$f_i(p_i; \boldsymbol{\theta}) = \frac{\phi(\frac{\ln p_i - \mu_p}{\sigma_p})}{\sigma_p(\Phi(\frac{\ln p^{\max} - \mu_p}{\sigma_p}) - \Phi(\frac{\ln \tilde{p} - \mu_p}{\sigma_p}))} \cdot \frac{1}{p_i}, \tag{3.27}$$

where ϕ and Φ are the pdf and cdf of the standard normal distribution, $\tilde{p} = \max\{p^{\min}, p_0\}$, and p_0 is the reservation productivity for an isolated worker satis-

[70] I use $\kappa = 10^{-6}$ in the estimation.

fying

$$\ln p_0 = \begin{cases} \ln b + \frac{\exp(\beta_0)}{1+\exp(\beta_0)}(\lambda_0 - \lambda_1) \int_{p_0}^{p^{max}} \frac{x^{-1} \cdot \frac{\Phi(\frac{\ln p^{\max}-\mu_p}{\sigma_p})-\Phi(\frac{x-\mu_p}{\sigma_p})}{\Phi(\frac{\ln p^{\max}-\mu_p}{\sigma_p})-\Phi(\frac{\ln p^{\min}-\mu_p}{\sigma_p})}}{\rho+\delta+\lambda_1 \frac{\exp(\beta_0)}{1+\exp(\beta_0)} \cdot \frac{\Phi(\frac{\ln p^{\max}-\mu_p}{\sigma_p})-\Phi(\frac{x-\mu_p}{\sigma_p})}{\Phi(\frac{\ln p^{\max}-\mu_p}{\sigma_p})-\Phi(\frac{\ln p^{\min}-\mu_p}{\sigma_p})}} dx & \text{if } \alpha = 1, \\ \\ \frac{1}{1-\alpha} \ln \left\{ b^{1-\alpha} + (1-\alpha) \frac{\exp(\beta_0)}{1+\exp(\beta_0)}(\lambda_0 - \lambda_1) \int_{p_0}^{p^{max}} \frac{x^{-\alpha} \cdot \frac{\Phi(\frac{\ln p^{\max}-\mu_p}{\sigma_p})-\Phi(\frac{x-\mu_p}{\sigma_p})}{\Phi(\frac{\ln p^{\max}-\mu_p}{\sigma_p})-\Phi(\frac{\ln p^{\min}-\mu_p}{\sigma_p})}}{\rho+\delta+\lambda_1 \frac{\exp(\beta_0)}{1+\exp(\beta_0)} \cdot \frac{\Phi(\frac{\ln p^{\max}-\mu_p}{\sigma_p})-\Phi(\frac{x-\mu_p}{\sigma_p})}{\Phi(\frac{\ln p^{\max}-\mu_p}{\sigma_p})-\Phi(\frac{\ln p^{\min}-\mu_p}{\sigma_p})}} dx \right\} & \text{if } \alpha \neq 1. \end{cases} \tag{3.28}$$

3.4 Additional Table for Model Fit

Moments	Data	Model
Co-movement in School Network		
Fraction of unemployed individuals experiencing UE transition		
...... if there is no school friend experiencing EU transition	0.1601	0.1612
...... if there are school friends experiencing EU transition	0.1631	0.1767
...... if there is no school friend experiencing EE transition	0.1603	0.1636
...... if there are school friends experiencing EE transition	0.1784	0.1922
Fraction of employed individuals experiencing EE transition		
...... if there is no school friend experiencing EU transition	0.0065	0.0063
...... if there are school friends experiencing EU transition	0.0086	0.0091
...... if there is no school friend experiencing EE transition	0.0071	0.0068
...... if there are school friends experiencing EE transition	0.0079	0.0128
Co-movement in Work Network		
Fraction of unemployed individuals experiencing UE transition		
...... if there is no work friend experiencing EU transition	0.0805	0.1381
...... if there are work friends experiencing EU transition	0.1766	0.1784
...... if there is no work friend experiencing EE transition	0.1547	0.1615
...... if there are work friends experiencing EE transition	0.1934	0.1920
Fraction of employed individuals experiencing EE transition		
...... if there is no work friend experiencing EU transition	0.0026	0.0041
...... if there are work friends experiencing EU transition	0.0080	0.0079
...... if there is no work friend experiencing EE transition	0.0055	0.0056
...... if there are work friends experiencing EE transition	0.0156	0.0175
Co-movement in Social-Activity Network		
Fraction of unemployed individuals experiencing UE transition		
...... if there is no social-activity friend experiencing EU transition	0.1624	0.1600
...... if there are social-activity friends experiencing EU transition	0.1547	0.1871
...... if there is no social-activity friend experiencing EE transition	0.1607	0.1629
...... if there are social-activity friends experiencing EE transition	0.1667	0.2051
Fraction of employed individuals experiencing EE transition		
...... if there is no social-activity friend experiencing EU transition	0.0072	0.0062
...... if there are social-activity friends experiencing EU transition	0.0071	0.0096
...... if there is no social-activity friend experiencing EE transition	0.0071	0.0067
...... if there are social-activity friends experiencing EE transition	0.0076	0.0100

Table 3.3: Model Fit: Moments in Job Transitions

Notes: A UE transition is defined as the transition from non-executive to executive. An EE transition is defined as the transition from one executive job to another with a productivity increase. An EU transition is defined as the transition from executive to non-executive.

3.5 Simulating Random Network with A Given Degree Sequence

To simulate a random network with a given degree sequence, I follow the Steger and Wormald (1999) algorithm.[71] The algorithm starts with an empty graph and adds edges sequentially. Each time, generate a potential edge by sampling uniformly a pair of nodes that have not yet received their full allotment of edges. Add this edge if it does not induce a loop or multiple edges. Continue this process until no more permissible pairs can be found. If the algorithm gets stuck in the sense that there are unmatched nodes left over that are not allowed to be paired with each other, start over and try again.

To simulate a series of monotonically growing (over time) networks, first generate the initial network exactly as described above. For the subsequent networks, to simulate Y^t, start the algorithm with Y^{t-1} instead of an empty graph and add edges as described above.

[71] It is a variant of the pairing model (also known as the configuration model or stubs model). See Blitzstein and Diaconis (2011) for a survey of other algorithms.

Chapter 4

Appendix for Chapter 2

4.1 Construction of Variable *industry**

We construct PCs' characteristic "*industry**" using the industrial breakdown information from `OpenSecrets.org`. Their breakdown has three nested levels - from coarse to fine - "sector", "industry", and "category". [72] We define variable *industry** to be the sector for PCs in sectors that are relatively homogeneous such as agricultural business. We define *industry** to be a finer level, industry or category, for PCs in sectors that are relatively heterogeneous such as miscellaneous business. The reason is that we use interaction term "same *industry**" in our analysis; and we want, within each *industry**, similar level of heterogeneity. A detailed description of the construction of the variable *industry** is given below, and the corresponding codebook is given in Table 4.1.

1. *industry**=sector if:

   A PC belongs to one of the following sectors: Agribusiness, Communications/Electronics, Construction, Defense, Energy & Natural Resources, Finance, Insurance & Real Estate, Health, Labor, Lawyers & Lobbyists, Transportation.

[72] The codebook is available at `https://www.opensecrets.org/downloads/crp/CRP_Categories.txt`.

2. *industry**=industry if:

    a) A PC belongs to the Miscellaneous Business sector.

    b) A PC belongs to the Other sector, but not in the Other industry.

    c) A PC belongs to the Ideological/Single-Issue sector, but not in one of the following industries: Misc Issues, Republican/Conservative, Democratic/Liberal, Leadership PACs, Candidate Committee.

3. *industry**=category if:

    a) A PC belongs to the Other industry in the Other sector, but not in the Other category.

    b) A PC belongs to the Misc Issues industry in the Ideological/Single-Issue sector.

4. *industry**=NA if:[73]

    a) A PC belongs to one of the following sectors: Unknown, Joint Candidate Committee, Party Committee, Candidate, Non-contribution.

    b) A PC belongs to one of the following industries in the Ideological/Single-Issue sector: Republican/Conservative, Democratic/Liberal, Leadership PACs, Candidate Committee.

    c) A PC's sector, industry and category are all Other.

[73]We define the interaction term "same *industry**" to be 0 if one or both PCs' *industry**s are NAs.

Code	Description
A	Agribusiness
B	Communic/Electronics
C	Construction
D	Defense
E	Energy/Nat Resource
F	Finance/Insur/RealEst
H	Health
H6000	Welfare & Social Work
J1300	Third-party committees
J3000	Consumer groups
J4000	Fiscal & tax policy
J7200	Elderly issues/Social Security
J7600	Animal Rights
J8000	Labor, anti-union
J9000	Other single-issue or ideological groups
K	Lawyers & Lobbyists
M	Transportation
N00	Business Associations
N01	Food & Beverage
N02	Beer, Wine & Liquor
N03	Retail Sales
N04	Misc Services
N05	Business Services
N06	Recreation/Live Entertainment
N07	Casinos/Gambling
N08	Lodging/Tourism
N12	Misc Business
N13	Chemical & Related Manufacturing
N14	Steel Production
N15	Misc Manufacturing & Distributing
N16	Textiles
NA	Not Available
P	Labor
Q04	Foreign & Defense Policy
Q05	Pro-Israel
Q08	Women's Issues
Q09	Human Rights
Q11	Environment
Q12	Gun Control
Q13	Gun Rights
Q14	Abortion Policy/Anti-Abortion
Q15	Abortion Policy/Pro-Abortion Rights
W02	Non-Profit Institutions
W03	Civil Servants/Public Officials
W04	Education
X5000	Military

Table 4.1: Codebook for Variable *industry**

4.2 Contribution Limit

Contribution limits for the 2003-2004 Election Cycle are given in Table 4.2.[74] They differ by the identity of the contributor and that of the recipient. As a contributor, individual, national party committee, non-national party committee, multicandidate PAC, and non-multicandidate PAC have different contribution limits. As a recipient, candidate or candidate committee, national party committee, non-national party committee, and other PC have different contribution limits. As shown in the table, there is no limit for contribution between party committees, national or local. For the rest of the contributions between PCs, the limits range from \$2,000 to \$25,000, most of which are set at \$5,000.

[74] Source: `http://www.fec.gov/pages/brochures/ContributionLimits2003-2004.htm`.

Contributor	Recipient				
	Candidate or candidate committee per election	National party committee per calendar year	State, district and local party committee per calendar year	Any other PC[1] per calendar year	Special Limits
Individual	$2,000	$25,000	$10000 (combined limit[6])	$5,000	$95,000 overall biennial limit: $37,500 to all candidates, $57,500 to all PACs and parties.[2]
National Party Committee	$5,000	No limit	No limit	$5,000	$35,000 to Senate candidate per campaign[3]
State, District & Local Party Committee	$5,000 (combined limit)	No limit	No limit	$5,000 (combined limit)	No limit
PAC (multicandidate[4])	$5,000	$15,000	$5,000 (combined limit)	$5,000	No limit
PAC (not multicandidate)	$2,000[5]	$25,000	$10,000 (combined limit)	$5,000	No limit

[1] A contribution earmarked for a candidate through a political committee counts against the original contributor's limit for that candidate. In certain circumstances, the contribution may also count against the contributor's limit to the PAC. 11 CFR 110.6. See also 11 CFR 110.1(h).
[2] No more than $37,500 of this amount may be contributed to state and local party committees and PACs.
[3] This limit is shared by the national committee and the Senate campaign committee.
[4] A multicandidate committee is a political committee with more than 50 contributors which has been registered for at least 6 months and, with the exception of state party committees, has made contributions to 5 or more candidates for federal office. 11 CFR 100.5(e)(3).
[5] A federal candidate's authorized committee(s) may contribute no more than $1,000 per election to another federal candidate's authorized committee(s). 11 CFR 102.12(c)(2).
[6] Combined limits are shared by federal accounts of all other state and local committees of the same party in the same state.

Table 4.2: Contribution Limits for the 2003-3004 Election Cycle

4.3 Figures and Tables: Monte Carlo I

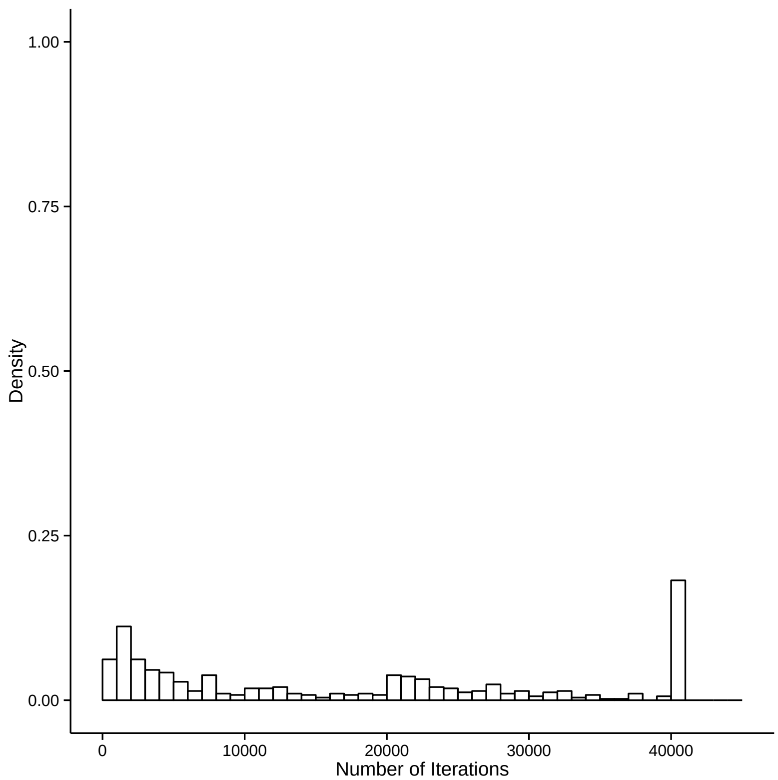

Figure 4.1: Histogram of Number of Iterations

Note: Bin size is 1000.

Average	Standard Deviation	Minimum	Maximum
0.011020	0.011637	0.000000	0.080000

Table 4.3: Summary Statistics on Misclassification Rates

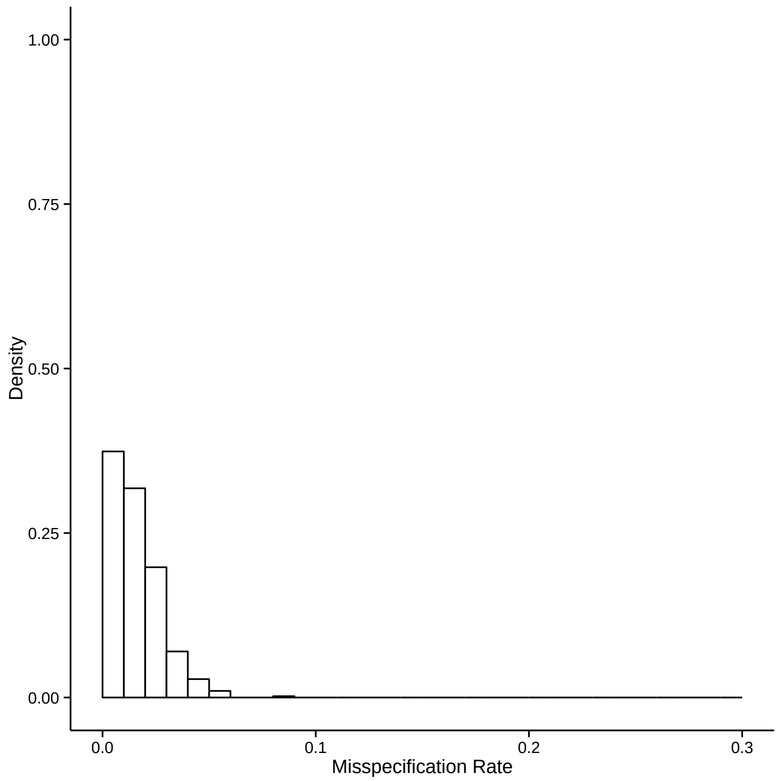

Figure 4.2: Histogram of Misclassification Rate

Note: Bin size is 0.01.

	Estimated Dem	Estimated Rep	Estimated Ind
True Dem	32.7520%	0.1820%	0.1540%
True Rep	0.1700%	32.7440%	0.2000%
True Ind	0.1840%	0.2120%	33.4020%

Table 4.4: Tabulation of Estimated vs. True Ideology

β	Bias	Standard Deviation
constant	-0.020032	0.306216
$\mathbb{1}_{x_i=x_j=\text{Dem}}$	0.014907	0.075353
$\mathbb{1}_{x_i=x_j=\text{Rep}}$	0.008465	0.075682
$\mathbb{1}_{x_i=x_j=\text{Ind}}$	0.017410	0.078703
$\mathbb{1}_{(x_i=\text{Dem},x_j=\text{Ind})\vee(x_i=\text{Ind},x_j=\text{Dem})}$	0.001838	0.064869
$\mathbb{1}_{(x_i=\text{Rep},x_j=\text{Ind})\vee(x_i=\text{Ind},x_j=\text{Rep})}$	-0.001370	0.064845
Same state	0.001803	0.043134
Same industry	0.002705	0.044662
One of them is a House campaign	-0.001016	0.046084
One of them is a Senate campaign	-0.004365	0.046909
One of them is a Presidential campaign	-0.001202	0.046442
One of them is a qualified PAC	0.000268	0.044655
One of them is a qualified Party	0.001389	0.042022
One of them is a national committee	0.001199	0.050153
One of them is authorized by a candidate	0.001038	0.046411
One of them is a joint fundraiser	-0.001212	0.051018
$(\ln b_i + \ln b_j)$	0.004666	0.103812
$((\ln b_i)^2 + (\ln b_j)^2)$	-0.000569	0.013868
$\ln b_i \ln b_j$	-0.000553	0.025041

Table 4.5: Parameters Governing Edge Formation Probabilities β

Note: Bias is defined as the difference between the average of posterior means across simulations and the true parameter value. Standard Deviation is defined as the standard deviation of posterior means across simulations.

θ	Bias	Standard Deviation
$\mathbb{P}(\text{Dem})$	-0.000223	0.004705
$\mathbb{P}(\text{Rep})$	-0.000224	0.004567
$\mathbb{P}(\text{Ind})$	0.000447	0.004620

Table 4.6: Parameters Governing the Fraction of Ideologies θ

Notes: Bias is defined as the difference between the average of posterior means across simulations and the true parameter value. Standard Deviation is defined as the standard deviation of posterior means across simulations.

$\mathbf{h}_{\text{Dem,Dem}}$	Bias	Standard Deviation
	0.136574	0.014174
	0.101386	0.011487
	-0.102415	0.014947
	-0.135545	0.017527
$\mathbf{h}_{\text{Rep,Rep}}$	Bias	Standard Deviation
	0.136643	0.014089
	0.102114	0.012182
	-0.102477	0.014942
	-0.136280	0.017799
$\mathbf{h}_{\text{Ind,Ind}}$	Bias	Standard Deviation
	0.134438	0.014390
	0.100568	0.011735
	-0.101278	0.014835
	-0.133727	0.017879
$\mathbf{h}_{\text{Dem,Rep}}$	Bias	Standard Deviation
	-0.097963	0.013732
	-0.033662	0.011047
	0.032859	0.009625
	0.098766	0.010588
$\mathbf{h}_{\text{Dem, Ind}}$	Bias	Standard Deviation
	-0.096669	0.013237
	-0.032747	0.011223
	0.032331	0.009789
	0.097085	0.009931
$\mathbf{h}_{\text{Rep, Ind}}$	Bias	Standard Deviation
	-0.097534	0.013724
	-0.032846	0.011521
	0.032789	0.009901
	0.097592	0.010524

Table 4.7: Parameters Governing the Weight Distribution h

Notes: Bias is defined as the difference between the average of posterior means across simulations and the true parameter value. Standard Deviation is defined as the standard deviation of posterior means across simulations.

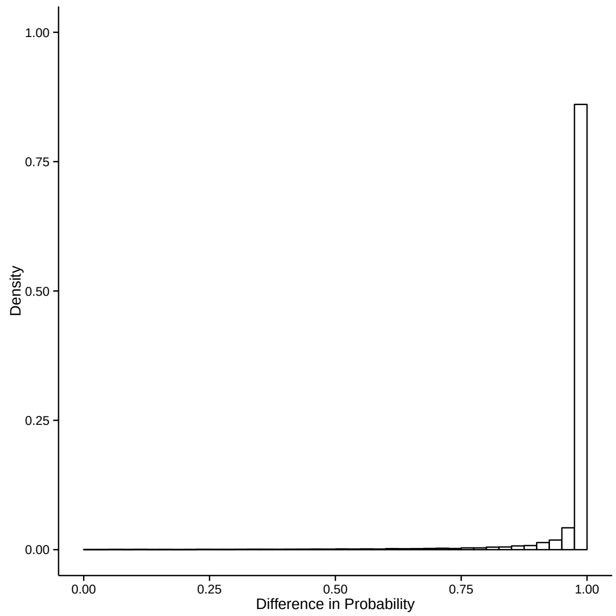

Figure 4.3: Histogram of Difference in Posterior Probability for Correctly Classified Vertices

Note: Each observation is a vertex. Horizontal axis is the difference between the highest posterior probability(i.e., the posterior probability of the true ideology) and the second highest posterior probability. Bin size is 0.025.

ϵ	Bias	Standard Deviation
	-0.002427	0.000660

Table 4.8: Parameter Governing Measurement Error ϵ

Notes: Bias is defined as the difference between the average of posterior means across simulations and the true parameter value. Standard Deviation is defined as the standard deviation of posterior means across simulations.

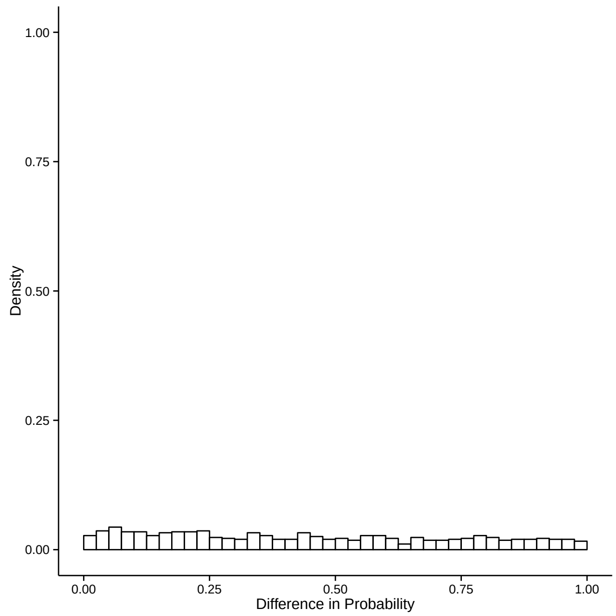

Figure 4.4: Histogram of Difference in Posterior Probability for Misclassified Vertices
Note: Each observation is a vertex. Horizontal axis is the difference between the highest posterior probability and the posterior probability of the true ideology. Bin size is 0.025.

4.4 Figures and Tables: Monte Carlo II

Average	Standard Deviation	Minimum	Maximum
0.056800	0.033604	0.000000	0.240000

Table 4.9: Summary Statistics on Misclassification Rates

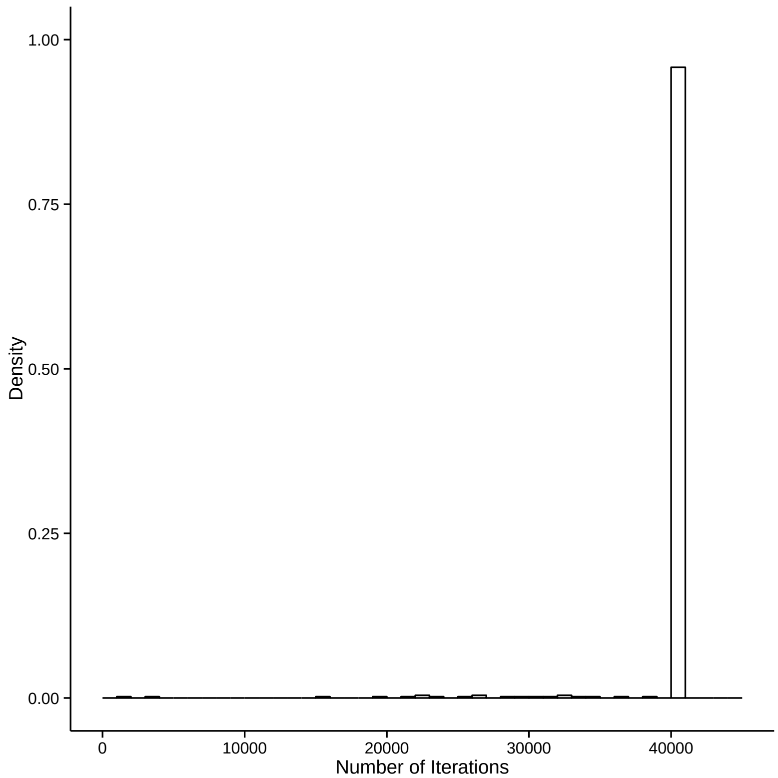

Figure 4.5: Histogram of Number of Iterations

Note: Bin size is 1000.

	Estimated Dem	Estimated Rep	Estimated Ind
True Dem	31.5080%	0.9440%	0.8760%
True Rep	1.0740%	31.3140%	0.8820%
True Ind	0.8940%	1.0100%	31.4980%

Table 4.10: Tabulation of Estimated vs. True Ideology

β	Bias	Standard Deviation
constant	-0.053234	0.278777
$\mathbb{1}_{x_i=x_j=\text{Dem}}$	0.004676	0.088396
$\mathbb{1}_{x_i=x_j=\text{Rep}}$	0.005889	0.085760
$\mathbb{1}_{x_i=x_j=\text{Ind}}$	0.008628	0.090062
$\mathbb{1}_{(x_i=\text{Dem},x_j=\text{Ind})\vee(x_i=\text{Ind},x_j=\text{Dem})}$	-0.005467	0.069264
$\mathbb{1}_{(x_i=\text{Rep},x_j=\text{Ind})\vee(x_i=\text{Ind},x_j=\text{Rep})}$	0.000659	0.071066
Same state	0.004227	0.043947
Same industry	0.002177	0.043028
One of them is a House campaign	0.001558	0.046981
One of them is a Senate campaign	0.000138	0.051027
One of them is a Presidential campaign	0.001651	0.046874
One of them is a qualified PAC	0.002929	0.048328
One of them is a qualified Party	0.002520	0.043062
One of them is a national committee	0.001504	0.048683
One of them is authorized by a candidate	0.005082	0.047008
One of them is a joint fundraiser	0.007028	0.049847
$(\ln b_i + \ln b_j)$	0.013339	0.095571
$((\ln b_i)^2 + (\ln b_j)^2)$	-0.001264	0.013323
$\ln b_i \ln b_j$	-0.002407	0.023131

Table 4.11: Parameters Governing Edge Formation Probabilities β

Note: Bias is defined as the difference between the average of posterior means across simulations and the true parameter value. Standard Deviation is defined as the standard deviation of posterior means across simulations.

θ	Bias	Standard Deviation
$\mathbb{P}$(Dem)	0.000043	0.004747
$\mathbb{P}$(Rep)	-0.000114	0.004533
$\mathbb{P}$(Ind)	0.000071	0.004761

Table 4.12: Parameters Governing the Fraction of Ideologies θ

Note: Bias is defined as the difference between the average of posterior means across simulations and the true parameter value. Standard Deviation is defined as the standard deviation of posterior means across simulations.

$\mathbf{h}_{\text{Dem,Dem}}$	Bias	Standard Deviation
	0.039407	0.010862
	0.070890	0.011031
	-0.070779	0.012694
	-0.039519	0.011386

$\mathbf{h}_{\text{Rep,Rep}}$	Bias	Standard Deviation
	0.038909	0.010678
	0.072241	0.011493
	-0.071994	0.012912
	-0.039157	0.012040

$\mathbf{h}_{\text{Ind,Ind}}$	Bias	Standard Deviation
	0.039131	0.011047
	0.071208	0.011610
	-0.070468	0.013130
	-0.039871	0.012104

$\mathbf{h}_{\text{Dem,Rep}}$	Bias	Standard Deviation
	-0.100952	0.014387
	-0.036391	0.011860
	0.036400	0.011109
	0.100943	0.010614

$\mathbf{h}_{\text{Dem,Ind}}$	Bias	Standard Deviation
	-0.100720	0.013911
	-0.035439	0.011370
	0.035755	0.009990
	0.100404	0.010030

$\mathbf{h}_{\text{Rep,Ind}}$	Bias	Standard Deviation
	-0.101129	0.014001
	-0.036941	0.011595
	0.036431	0.010460
	0.101638	0.010507

Table 4.13: Parameters Governing the Weight Distribution h

Note: Bias is defined as the difference between the average of posterior means across simulations and the true parameter value. Standard Deviation is defined as the standard deviation of posterior means across simulations.

ϵ	Bias	Standard Deviation
	-0.002484	0.000581

Table 4.14: Parameter Governing Measurement Error ϵ

Note: Bias is defined as the difference between the average of posterior means across simulations and the true parameter value. Standard Deviation is defined as the standard deviation of posterior means across simulations.

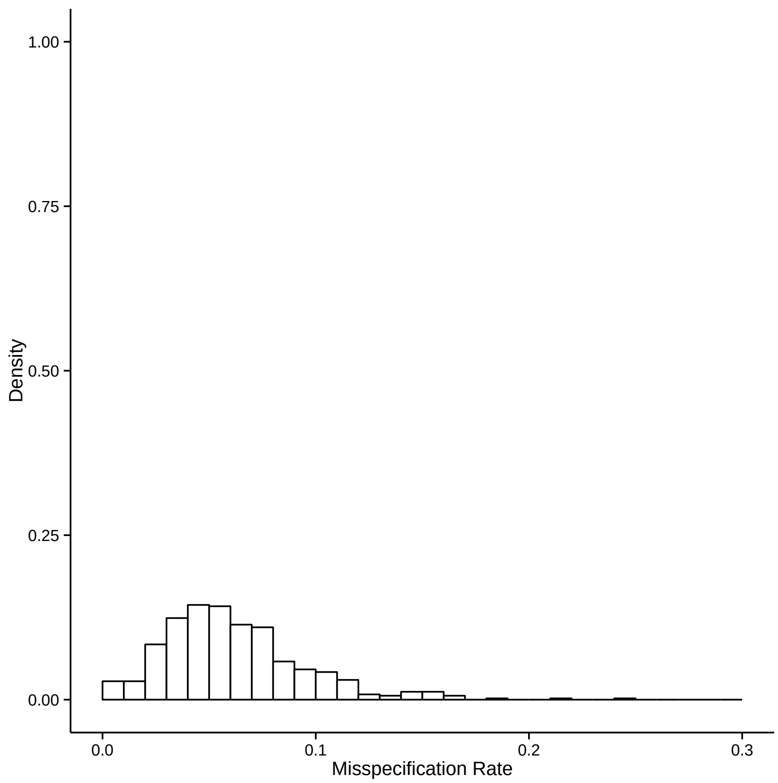

Figure 4.6: Histogram of Misclassification Rate

Note: Bin size is 0.01.

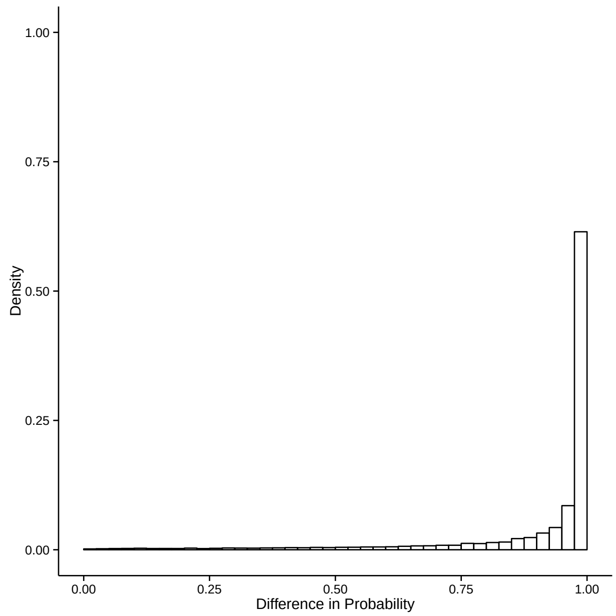

Figure 4.7: Histogram of Difference in Posterior Probability for Correctly Classified Vertices

Note: Each observation is a vertex. Horizontal axis is the difference between the highest posterior probability(i.e., the posterior probability of the true ideology) and the second highest posterior probability. Bin size is 0.025.

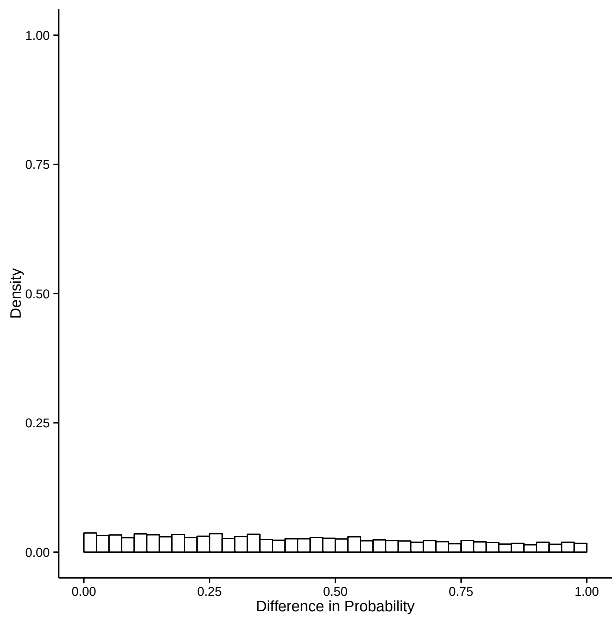

Figure 4.8: Histogram of Difference in Posterior Probability for Misclassified Vertices

Note: Each observation is a vertex. Horizontal axis is the difference between the highest posterior probability and the posterior probability of the true ideology. Bin size is 0.025.

4.5 Figures and Tables: Monte Carlo III

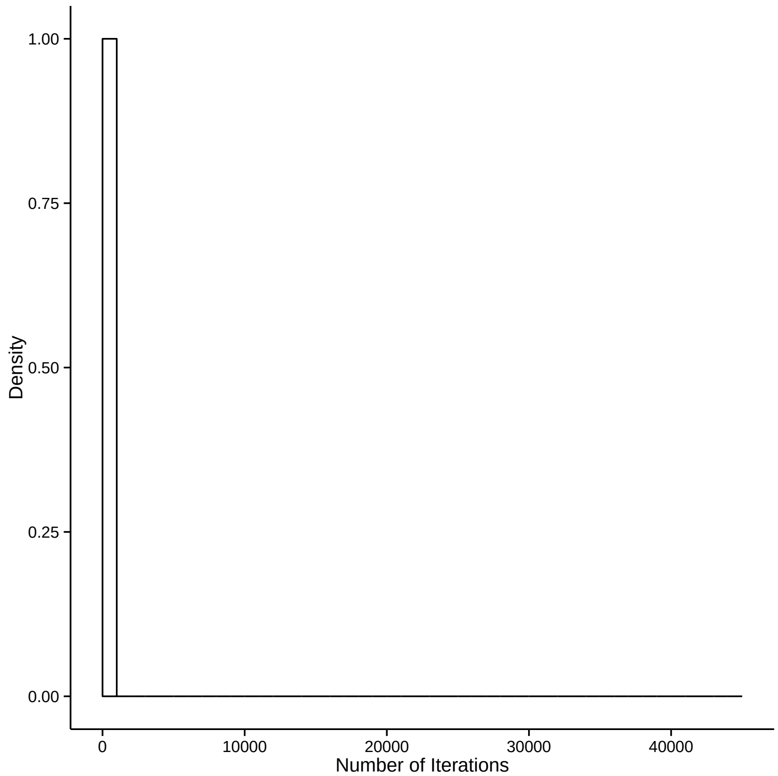

Figure 4.9: Histogram of Number of Iterations

Note: Bin size is 1000.

β	Bias	Standard Deviation
constant	-0.000600	0.045500
$\mathbb{1}_{x_i=x_j=\text{Dem}}$	0.000375	0.013841
$\mathbb{1}_{x_i=x_j=\text{Rep}}$	0.001279	0.014102
$\mathbb{1}_{x_i=x_j=\text{Ind}}$	0.000725	0.014023
$\mathbb{1}_{(x_i=\text{Dem},x_j=\text{Ind})\vee(x_i=\text{Ind},x_j=\text{Dem})}$	0.001145	0.012192
$\mathbb{1}_{(x_i=\text{Rep},x_j=\text{Ind})\vee(x_i=\text{Ind},x_j=\text{Rep})}$	0.000070	0.011883
Same state	-0.000222	0.008429
Same industry	-0.000045	0.008724
One of them is a House campaign	-0.000023	0.009036
One of them is a Senate campaign	-0.000173	0.009085
One of them is a Presidential campaign	-0.000214	0.009082
One of them is a qualified PAC	-0.000737	0.008557
One of them is a qualified Party	-0.000620	0.008545
One of them is a national committee	0.000057	0.009115
One of them is authorized by a candidate	-0.000122	0.009020
One of them is a joint fundraiser	0.000534	0.009923
$(\ln b_i + \ln b_j)$	0.000375	0.015589
$((\ln b_i)^2 + (\ln b_j)^2)$	-0.000113	0.002002
$\ln b_i \ln b_j$	0.000054	0.004421

Table 4.15: Parameters Governing Edge Formation Probabilities β

Note: Bias is defined as the difference between the average of posterior means across simulations and the true parameter value. Standard Deviation is defined as the standard deviation of posterior means across simulations.

θ	Bias	Standard Deviation
$\mathbb{P}$(Dem)	0.000469	0.007948
$\mathbb{P}$(Rep)	-0.000514	0.007469
$\mathbb{P}$(Ind)	0.000046	0.007534

Table 4.16: Parameters Governing the Fraction of Ideologies θ

Note: Bias is defined as the difference between the average of posterior means across simulations and the true parameter value. Standard Deviation is defined as the standard deviation of posterior means across simulations.

$\mathbf{h}_{\text{Dem,Dem}}$	Bias	Standard Deviation
	0.003591	0.005047
	0.007005	0.004737
	-0.006713	0.006200
	-0.003883	0.006027

$\mathbf{h}_{\text{Rep,Rep}}$	Bias	Standard Deviation
	0.003633	0.005297
	0.007228	0.004696
	-0.007282	0.006413
	-0.003580	0.005937

$\mathbf{h}_{\text{Ind,Ind}}$	Bias	Standard Deviation
	0.003769	0.005023
	0.007100	0.004713
	-0.007326	0.005899
	-0.003543	0.005966

$\mathbf{h}_{\text{Dem, Rep}}$	Bias	Standard Deviation
	-0.009173	0.006111
	-0.003547	0.005576
	0.002779	0.004879
	0.009941	0.003743

$\mathbf{h}_{\text{Dem,Ind}}$	Bias	Standard Deviation
	-0.009563	0.005808
	-0.003075	0.005394
	0.003214	0.004920
	0.009424	0.003579

$\mathbf{h}_{\text{Rep, Ind}}$	Bias	Standard Deviation
	-0.009567	0.005856
	-0.003141	0.005524
	0.003000	0.004760
	0.009708	0.003814

Table 4.17: Parameters Governing the Weight Distribution $\mathbf{h}$

Note: Bias is defined as the difference between the average of posterior means across simulations and the true parameter value. Standard Deviation is defined as the standard deviation of posterior means across simulations.

ϵ	Bias	Standard Deviation
	-0.002183	0.001299

Table 4.18: Parameter Governing Measurement Error ϵ

Note: Bias is defined as the difference between the average of posterior means across simulations and the true parameter value. Standard Deviation is defined as the standard deviation of posterior means across simulations.

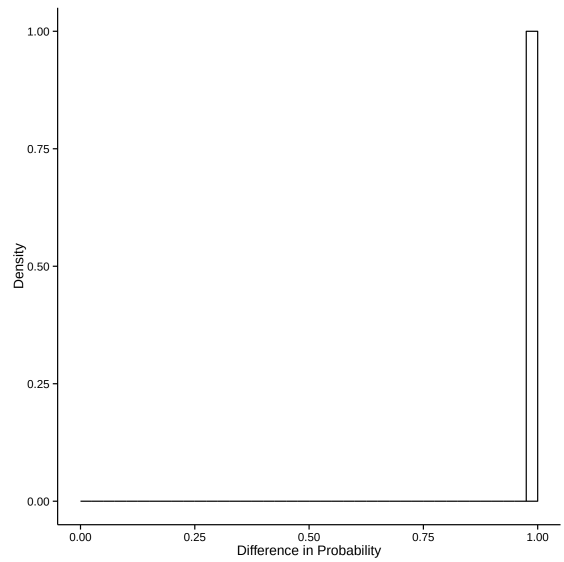

Figure 4.10: Histogram of Difference in Posterior Probability for Correctly Classified Vertices

Note: Each observation is a vertex. Horizontal axis is the difference between the highest posterior probability(i.e., the posterior probability of the true ideology) and the second highest posterior probability. Bin size is 0.025.

4.6 Figures and Tables: Monte Carlo IV

β	Bias	Standard Deviation
constant	-0.000099	0.003286
$\mathbb{1}_{x_i=x_j=\text{Dem}}$	0.000333	0.002225
$\mathbb{1}_{x_i=x_j=\text{Rep}}$	0.000288	0.002065
$\mathbb{1}_{x_i=x_j=\text{Ind}}$	0.000045	0.002128
$\mathbb{1}_{(x_i=\text{Dem},x_j=\text{Ind})\vee(x_i=\text{Ind},x_j=\text{Dem})}$	0.000012	0.002087
$\mathbb{1}_{(x_i=\text{Rep},x_j=\text{Ind})\vee(x_i=\text{Ind},x_j=\text{Rep})}$	0.000441	0.002084
Same state	0.000135	0.001168
Same industry	-0.000033	0.001090
One of them is a House campaign	-0.000167	0.001289
One of them is a Senate campaign	-0.000063	0.001331
One of them is a Presidential campaign	-0.000026	0.001187
One of them is a qualified PAC	0.000091	0.001104
One of them is a qualified Party	0.000060	0.001147
One of them is a national committee	0.000015	0.001267
One of them is authorized by a candidate	-0.000166	0.001376
One of them is a joint fundraiser	0.000039	0.001281
$(\ln b_i + \ln b_j)$	-0.000084	0.000527
$((\ln b_i)^2 + (\ln b_j)^2)$	0.000005	0.000216
$\ln b_i \ln b_j$	0.000030	0.000583

Table 4.19: Parameters Governing Edge Formation Probabilities β

Note: Bias is defined as the difference between the average of posterior means across simulations and the true parameter value. Standard Deviation is defined as the standard deviation of posterior means across simulations.

θ	Bias	Standard Deviation
$\mathbb{P}$(Dem)	-0.000759	0.006007
$\mathbb{P}$(Rep)	0.000469	0.004874
$\mathbb{P}$(Ind)	0.000289	0.005318

Table 4.20: Parameters Governing the Fraction of Ideologies θ

Note: Bias is defined as the difference between the average of posterior means across simulations and the true parameter value. Standard Deviation is defined as the standard deviation of posterior means across simulations.

$\mathbf{h}_{\text{Dem,Dem}}$	Bias	Standard Deviation
	0.000352	0.000592
	0.000110	0.000951
	0.000025	0.000896
	-0.000488	0.001082

$\mathbf{h}_{\text{Rep,Rep}}$	Bias	Standard Deviation
	0.000309	0.000557
	-0.000051	0.000850
	0.000283	0.000831
	-0.000542	0.001025

$\mathbf{h}_{\text{Ind,Ind}}$	Bias	Standard Deviation
	-0.000002	0.001127
	0.000030	0.001127
	0.000158	0.001154
	-0.000186	0.001181

$\mathbf{h}_{\text{Dem,Rep}}$	Bias	Standard Deviation
	-0.001770	0.001944
	0.000260	0.001699
	0.000455	0.001516
	0.001056	0.001108

$\mathbf{h}_{\text{Dem,Ind}}$	Bias	Standard Deviation
	-0.000148	0.001221
	-0.000079	0.001344
	-0.000163	0.001343
	0.000389	0.000966

$\mathbf{h}_{\text{Rep,Ind}}$	Bias	Standard Deviation
	-0.000176	0.001315
	-0.000162	0.001176
	-0.000244	0.001361
	0.000581	0.000852

Table 4.21: Parameters Governing the Weight Distribution $\mathbf{h}$

Note: Bias is defined as the difference between the average of posterior means across simulations and the true parameter value. Standard Deviation is defined as the standard deviation of posterior means across simulations.

ϵ	Bias	Standard Deviation
	-0.001132	0.002546

Table 4.22: Parameter Governing Measurement Error ϵ

Note: Bias is defined as the difference between the average of posterior means across simulations and the true parameter value. Standard Deviation is defined as the standard deviation of posterior means across simulations.

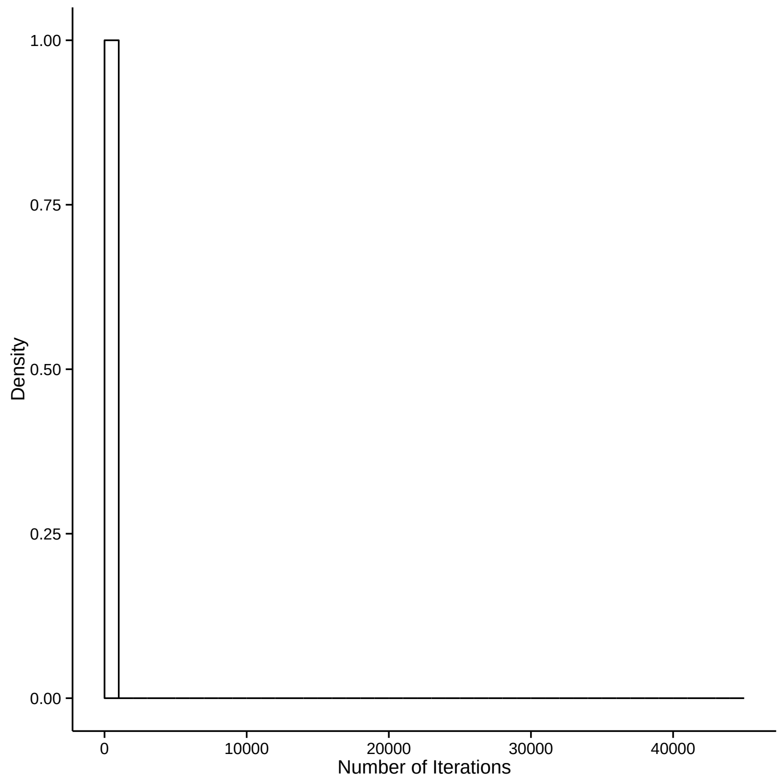

Figure 4.11: Histogram of Number of Iterations

Note: Bin size is 1000.

4.7 List of PCs with Estimated Ideology Different from Self Report

FEC ID	Committee Name
C00113662	NEW MEXICANS FOR BILL RICHARDSON
C00165753	LEADERSHIP '02 (FKA FRIENDS OF ALBERT GORE JR INC)
C00247734	COMMITTEE TO ELECT WILLIAM J JEFFERSON TO THE UNITED STATES CONGRESS
C00401364	FRIENDS OF JOHN SWEENEY
C00178418	BOUCHER FOR CONGRESS COMMITTEE
C00387829	DEMOCRAT GRAYSON FOR THE HOUSE
C00396101	JON PORTER FOR CONGRESS COMMITTEE
C00316596	CHRIS JOHN FOR CONGRESS
C00316141	RE-ELECT HAROLD FORD
C00220145	GENE TAYLOR FOR CONGRESS COMMITTEE
C00223230	FRIENDS OF JOHN TANNER
C00366401	MARK PRYOR FOR US SENATE
C00315176	FEINSTEIN FOR SENATE
C00347310	FRIENDS OF CHRIS DODD 2004
C00386292	NORTH DAKOTA 2004
C00143438	FRIENDS OF BYRON DORGAN
C00305110	A LOT OF PEOPLE WHO SUPPORT JEFF BINGAMAN
C00364364	DANIEL K INOUYE IN 2004
C00280917	DANIEL K INOUYE IN 98
C00396044	JOHN KENNEDY FOR US SENATE INC
C00391110	VICTORY 2004
C00389957	COBLE FOR US SENATE
C00224972	FRIENDS OF SENATOR ROCKEFELLER
C00215830	JOHN BREAUX COMMITTEE
C00391862	LOUISIANA SENATE 2003
C00325126	FRIENDS OF MARY LANDRIEU INC
C00317214	MARY LANDRIEU FOR SENATE COMMITTEE INC
C00202754	FRIENDS OF KENT CONRAD
C00368209	NELSON 2006
C00306712	NELSON 2000
C00385013	NEVADA SENATE 2004
C00204370	FRIENDS FOR HARRY REID
C00387449	MONTANA NEVADA VICTORY FUND
C00308676	WYDEN FOR SENATE
C00201533	TIM JOHNSON FOR SOUTH DAKOTA INC
C00402008	MONTANA ARKANSAS VICTORY FUND
C00255463	FRIENDS OF BLANCHE LINCOLN
C00385633	ARKANSAS SENATE 2004
C00349217	CARPER FOR SENATE
C00344051	BILL NELSON FOR U S SENATE
C00306860	EVAN BAYH COMMITTEE
C00383497	BAUCUS VICTORY FUND
C00328211	FRIENDS OF MAX BAUCUS 2002

Table 4.23: PCs that Self Reported to be Democratic, but are Estimated to be Republican

FEC ID	Committee Name
C00400598	ILLINOIS US SENATE VICTORY COMMITTEE
C00387001	OHIO VICTORY COMMITTEE
C00406322	REPUBLICAN PARTY OF KENDALL COUNTY
C00188078	SEVENTH CONGRESSIONAL DISTRICT REPUBLICAN PARTY OF WISCONSIN
C00350496	BRADY FOR CONGRESS
C00371443	DANNY DAVIS FOR CONGRESS
C00400531	KY 04 CONGRESSIONAL VICTORY COMMITTEE
C00376749	RODNEY ALEXANDER FOR CONGRESS INC.
C00272211	PETE KING FOR CONGRESS COMMITTEE
C00272153	COMMITTEE TO ELECT MCHUGH
C00378158	ZARELLI FOR CONGRESS
C00396523	ROSELYN FOR CONGRESS
C00401703	FRIENDS OF ALJANICH
C00385542	PHELPS FOR CONGRESS
C00388884	HOOSIERS FOR HARDY
C00400556	LA 03 CONGRESSIONAL VICTORY COMMITTEE
C00400507	LA 07 CONGRESSIONAL VICTORY COMMITTEE
C00190637	TIERNEY FOR CONGRESS COMMITTEE
C00386060	DEROSSETT FOR CONGRESS
C00399675	FRIENDS OF STEVE MORROW
C00392860	BRAUNER FOR CONGRESS
C00398776	HUFFMAN FOR CONGRESS
C00403642	PAUL RODRIGUEZ FOR CONGRESS
C00386078	BELL FOR CONGRESS COMMITTEE
C00388991	RICCARDI FOR CONGRESS
C00399295	MATT MUEDA FOR CONGRESS
C00395061	JANE ESHAGPOOR FOR CONGESS
C00398834	ASAY FOR CONGRESS COMMITTEE
C00331108	REP DON YOUNG CONSTITUTIONAL DEFENSE FUND
C00320168	ASA HUTCHINSON FOR CONGRESS COMMITTEE
C00395731	TERESA DOGGETT TAYLOR FOR CONGRESS
C00333294	DOUG OSE FOR CONGRESS '98
C00219204	PORTER GOSS RE-ELECTION TEAM
C00335190	CONNELLY FOR CONGRESS
C00300699	NODLER FOR CONGRESS COMMITTEE
C00096412	COMMITTEE TO REELECT CONGRESSMAN CHRIS SMITH
C00091298	THE COMMITTEE TO RE-ELECT CONGRESSWOMAN MARGE ROUKEMA
C00334334	DON SHERWOOD FOR CONGRESS
C00397075	SANTA CRUZ ACTION COMMITTEE
C00394346	BUSH ADMINISTRATION RETIREMENT FUND PAC (BARF PAC)
C00405688	DUMP BUSH MISSOULA
C00374652	DIANE ALLEN FOR US SENATE
C00389692	DR KATHURIA FOR US SENATE
C00349795	GORMLEY FOR SENATE PRIMARY ELECTION FUND
C00366237	CHAFEE FOR SENATE
C00325571	SENATOR JOHN WARNER COMMITTEE

Table 4.24: PCs that Self Reported to be Republican, but are Estimated to be Democratic

4.8 Alternative Model: No Measurement Error, but with Hold Out Sample

In this alternative model, we randomly select 200 PCs in $\mathcal{V}^o$ to be the holdout sample. Additionally, we assume no measurement error, i.e., $\epsilon = 0$ and $x_i = \hat{x}_i$ when $\hat{x}_i$ is available. We estimate this model pretending $\hat{x}_i$ is not available for the holdout sample. In order to assess our estimates, we compare the estimated posterior distribution and the self report for the holdout sample.

Posterior distributions of $\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{h}$ are summarized in Table 4.25, 4.26, and Figure 4.12.

$\boldsymbol{\beta}$	Posterior Mean	Posterior Standard Deviation
constant	-72.9487	1.1780
$\mathbb{1}_{x_i=x_j=\text{Dem}}$	47.9304	0.4056
$\mathbb{1}_{x_i=x_j=\text{Rep}}$	47.8478	0.4027
$\mathbb{1}_{x_i=x_j=\text{Ind}}$	47.8889	0.3907
$\mathbb{1}_{(x_i=\text{Dem},x_j=\text{Ind})\vee(x_i=\text{Ind},x_j=\text{Dem})}$	48.3420	0.4055
$\mathbb{1}_{(x_i=\text{Rep},x_j=\text{Ind})\vee(x_i=\text{Ind},x_j=\text{Rep})}$	48.4466	0.4062
Same state	0.7792	0.0114
Same industry	0.5151	0.0272
One of them is a House campaign	0.5315	0.0082
One of them is a Senate campaign	0.2261	0.0044
One of them is a Presidential campaign	-0.2852	0.0169
One of them is a qualified PAC	0.3990	0.0051
One of them is a qualified Party	-0.7151	0.0133
One of them is a national committee	0.7163	0.0116
One of them is authorized by a candidate	-0.4360	0.0071
One of them is a joint fundraiser	-0.8477	0.0102
$(\ln b_i + \ln b_j)$	1.4648	0.0579
$((\ln b_i)^2 + (\ln b_j)^2)$	0.0056	0.0003
$\ln b_i \ln b_j$	-0.1017	0.0038

Table 4.25: Alternative Model: Posterior Distribution of β

In the following part, we focus on the holdout sample. Table 4.27 cross tabulates

$\boldsymbol{\theta}$	Posterior Mean	Posterior Standard Deviation
$\mathbb{P}$(Dem)	0.3839	0.0045
$\mathbb{P}$(Rep)	0.3913	0.0045
$\mathbb{P}$(Ind)	0.2248	0.0038

Table 4.26: Alternative Model: Posterior Distribution of θ

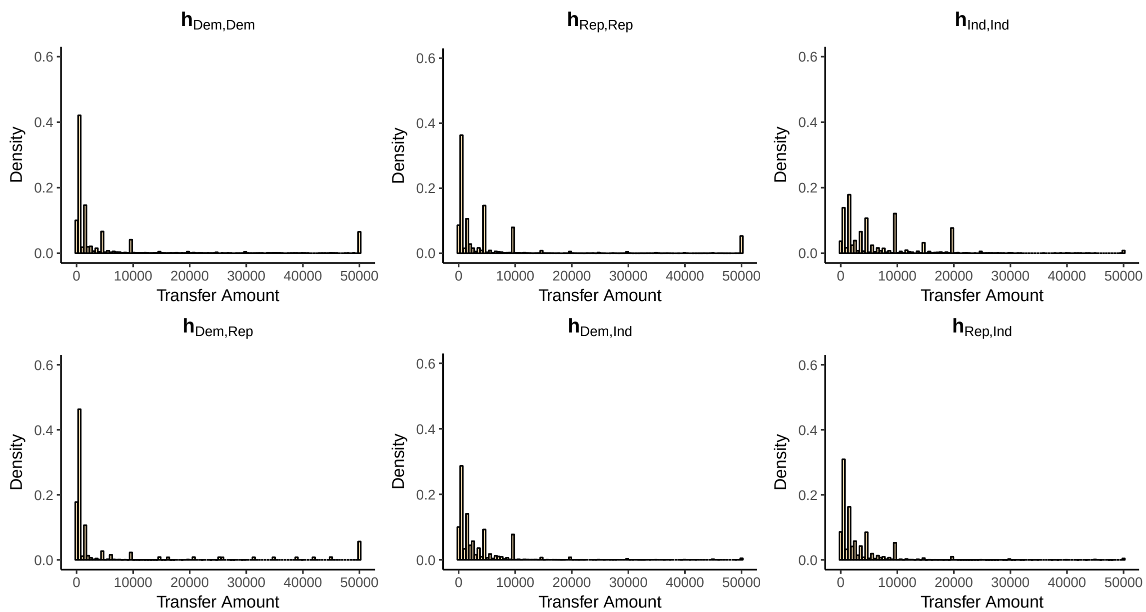

Figure 4.12: Posterior Mean of h
Note: Probability of transfer amount higher than $50,000 is plotted at $50,000.

PCs in the holdout sample according to self-reported and estimated ideology. Our estimated ideology matches the self report for 87.76% of those who reported to be Democratic and 79.17% of those who reported to be Republican.

In this part, we analyze the posterior distribution of political ideology. Figure 4.13 plots, for PCs whose estimates are the same as the self-reports (i.e., the reported ideology has the highest posterior probability), the distribution of the differences between the highest and the second highest posterior probability. These differences concentrate around 1, meaning the posterior probabilities concentrate on the self-reported affiliation. This confirms that we do not obtain these classifications by luck.

We do a similar analysis in Figure 4.14, for PCs whose estimates differ from the self-reports, by plotting the distributions of the differences between the highest posterior probability and the posterior probability of the self-reported ideology. The distributions for PCs which self reported to be Democratic (Republican) but are estimated as Republican (Democratic) are concentrated around 0, indicating that these are "near misses".

Finally, Table 4.28 and 4.29 list the PCs that self reported to be Democratic (Republican), but are estimated to be Republican (Democratic).

	Estimated Dem	Estimated Rep	Estimated Ind
Self-Reported Dem	86 (87.76%)	9 (9.18%)	3 (3.06%)
Self-Reported Rep	10 (10.42%)	76 (79.17%)	10 (10.42%)
Self-Reported Ind	1 (16.67%)	3 (50.00%)	2 (33.33%)

Table 4.27: Tabulation of Estimated vs. Self-Reported Ideology
Note: The percentages are calculated for each row.

FEC ID	Committee Name
C00327403	FRIENDS OF JONATHAN MILLER
C00367060	JOHN MILKOVICH FOR CONGRESS
C00381350	MARK BUDETICH
C00388454	JOHNSON FOR US SENATE
C00390245	REYES FOR CONGRESS
C00394858	VICTORY 04
C00399097	JOHN SALAZAR AND KEN SALAZAR JOINT COMMITTEE
C00399154	BURKS FOR US SENATE CAMPAIGN COMMITTEE
C00402149	FRIENDS TO ELECT JEFF MILLER

Table 4.28: PCs that Self Reported to be Democratic, but are Estimated to be Republican

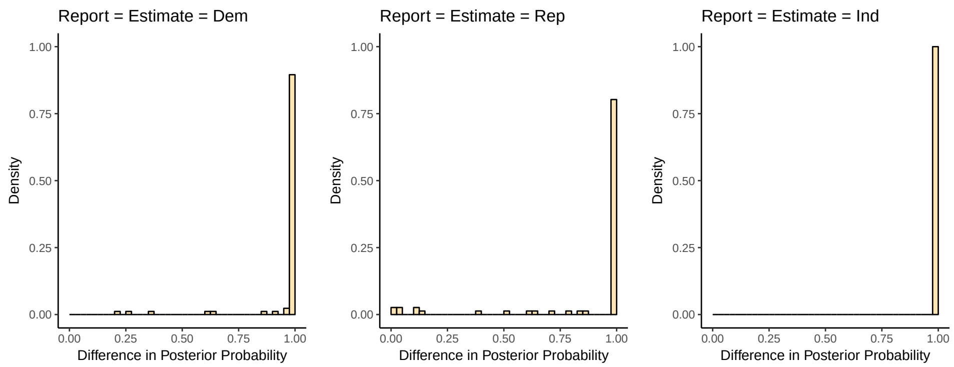

Figure 4.13: Distribution of Difference in the Posterior Probability of Ideology
Note: Horizontal axis is the difference between the highest posterior probability (i.e., the posterior probability of the self-reported ideology) and the second highest posterior probability. Bin size is 0.025.

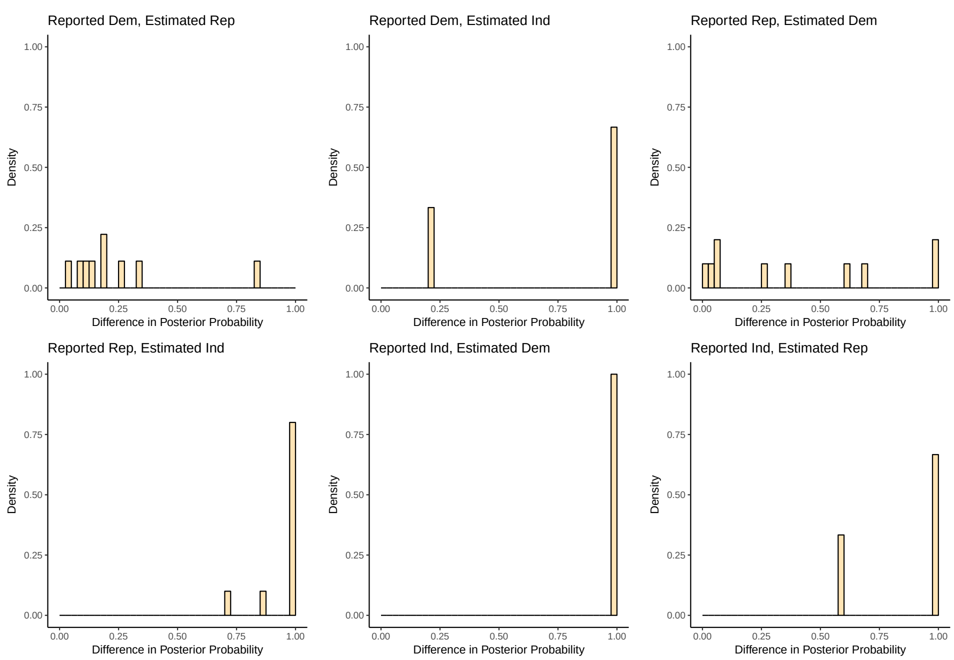

Figure 4.14: Distribution of Difference in the Posterior Probability of Ideology
Note: Horizontal axis is the difference between the highest posterior probability and the posterior probability of the self-reported ideology. Bin size is 0.025.

FEC ID	Committee Name
C00188078	SEVENTH CONGRESSIONAL DISTRICT REPUBLICAN PARTY OF WISCONSIN
C00349795	GORMLEY FOR SENATE PRIMARY ELECTION FUND
C00367839	SALAZAR FOR CONGRESS
C00371443	DANNY DAVIS FOR CONGRESS
C00375485	RUSTY GLOVER FOR CONGRESS
C00386078	BELL FOR CONGRESS COMMITTEE
C00387571	STARK FOR CONGRESS
C00389130	FRIENDS OF JOE NEGRON
C00390203	RISLEY FOR CONGRESS
C00404772	RANDY EASTWOOD FOR CONGRESS

Table 4.29: PCs that Self Reported to be Republican, but are Estimated to be Democratic

Bibliography

Abbe, E. (2017). Community detection and stochastic block models: Recent developments. *arXiv preprint arXiv:1703.10146*.

Abbe, E., A. S. Bandeira, and G. Hall (2014). Exact recovery in the stochastic block model. *arXiv preprint arXiv:1405.3267*.

Abbe, E. and C. Sandon (2015). Community detection in general stochastic block models: Fundamental limits and efficient algorithms for recovery. In *Foundations of Computer Science (FOCS), 2015 IEEE 56th Annual Symposium on*, pp. 670–688. IEEE.

Addison, J. T. and P. Portugal (2002). Job search methods and outcomes. *Oxford Economic Papers 54*(3), 505–533.

Aizawa, N. and H. Fang (2018). Equilibrium labor market search and health insurance reform. Technical report, National Bureau of Economic Research.

Albert, J. H. and S. Chib (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American statistical Association 88*(422), 669–679.

Altonji, J. G. and L. M. Segal (1996). Small-sample bias in gmm estimation of covariance structures. *Journal of Business & Economic Statistics 14*(3), 353–366.

Arbex, M., D. O'Dea, and D. Wiczer (2018). Network search: Climbing the job ladder faster. *International Economic Review*.

Arrow, K. J. and R. Borzekowski (2004). Limited network connections and the distribution of wages.

Bagger, J., F. Fontaine, F. Postel-Vinay, and J.-M. Robin (2014). Tenure, experience, human capital, and wages: A tractable equilibrium search model of wage dynamics. *American Economic Review 104*(6), 1551–96.

Banerjee, A. V., A. G. Chandrasekhar, E. Duflo, and M. O. Jackson (2017). Using gossips to spread information: Theory and evidence from two randomized controlled trials.

Barberá, P. (2015). Birds of the same feather tweet together: Bayesian ideal point estimation using twitter data. *Political Analysis 23*(1), 76–91.

Bebchuk, L. and J. Fried (2004). *Pay without performance*, Volume 29. Cambridge, MA: Harvard University Press.

Blitzstein, J. and P. Diaconis (2011). A sequential importance sampling algorithm for generating random graphs with prescribed degrees. *Internet mathematics 6*(4), 489–522.

Bonica, A. (2013). Ideology and interests in the political marketplace. *American Journal of Political Science 57*(2), 294–311.

Bontemps, C., J.-M. Robin, and G. J. Van den Berg (1999). An empirical equilibrium job search model with search on the job and heterogeneous workers and firms. *International Economic Review 40*(4), 1039–1074.

Bontemps, C., J.-M. Robin, and G. J. Van den Berg (2000). Equilibrium search with continuous productivity dispersion: Theory and nonparametric estimation. *International Economic Review 41*(2), 305–358.

Boorman, S. A. (1975). A combinatiorial optimization model for transmission of job information through contact networks. *The bell journal of economics*, 216–249.

Breiger, R. L., S. A. Boorman, and P. Arabie (1975). An algorithm for clustering relational data with applications to social network analysis and comparison with multidimensional scaling. *Journal of mathematical psychology 12*(3), 328–383.

Brown, M., E. Setren, and G. Topa (2016). Do informal referrals lead to better matches? evidence from a firm's employee referral system. *Journal of Labor Economics 34*(1), 161–209.

Burdett, K. and D. T. Mortensen (1998). Wage differentials, employer size, and unemployment. *International Economic Review*, 257–273.

Burks, S. V., B. Cowgill, M. Hoffman, and M. Housman (2015). The value of hiring through employee referrals. *The Quarterly Journal of Economics 130*(2), 805–839.

Cahuc, P., F. Postel-Vinay, and J.-M. Robin (2006). Wage bargaining with on-the-job search: Theory and evidence. *Econometrica 74*(2), 323–364.

Cai, T. T., T. Liang, and A. Rakhlin (2017). Weighted message passing and minimum energy flow for heterogeneous stochastic block models with side information. *arXiv preprint arXiv:1709.03907*.

Cai, Y. and M. Sevilir (2012). Board connections and m&a transactions. *Journal of Financial Economics 103*(2), 327–349.

Calvó-Armengol, A. and M. O. Jackson (2004). The effects of social networks on employment and inequality. *The American Economic Review 94*(3), 426–454.

Calvó-Armengol, A. and M. O. Jackson (2007). Networks in labor markets: Wage and employment dynamics and inequality. *Journal of economic theory 132*(1), 27–46.

Calvó-Armengol, A. and Y. Zenou (2005). Job matching, social network and word-of-mouth communication. *Journal of urban economics 57*(3), 500–522.

Chandrasekhar, A. and R. Lewis (2011). Econometrics of sampled networks. *Unpublished manuscript, MIT*.

Conley, T. G. and G. Topa (2007). Estimating dynamic local interactions models. *Journal of Econometrics 140*(1), 282–303.

Corcoran, M., L. Datcher, and G. Duncan (1980). Information and influence networks in labor markets. *Five thousand American families: Patterns of economic progress 8*(S 1), 37.

Dey, M. S. and C. J. Flinn (2005). An equilibrium model of health insurance provision and wage determination. *Econometrica 73*(2), 571–627.

Dustmann, C., A. Glitz, U. Schönberg, and H. Brücker (2015). Referral-based job search networks. *The Review of Economic Studies 83*(2), 514–546.

Eckstein, Z. and G. J. Van den Berg (2007). Empirical labor search: A survey. *Journal of Econometrics 136*(2), 531–564.

Edmans, A. and X. Gabaix (2016). Executive compensation: A modern primer. *Journal of Economic literature 54*(4), 1232–87.

El-Khatib, R., K. Fogel, and T. Jandik (2015). Ceo network centrality and merger performance. *Journal of Financial Economics 116*(2), 349–382.

Engelberg, J., P. Gao, and C. A. Parsons (2013). The price of a ceo's rolodex. *The Review of Financial Studies 26*(1), 79–114.

Fernandez, R. M., E. J. Castilla, and P. Moore (2000). Social capital at work: Networks and employment at a phone center. *American journal of sociology 105*(5), 1288–1356.

Flinn, C. and J. Heckman (1982). New methods for analyzing structural models of labor force dynamics. *Journal of Econometrics 18*(1), 115–168.

Fontaine, F. (2008). Why are similar workers paid differently? the role of social networks. *Journal of Economic Dynamics and Control 32*(12), 3960–3977.

Fracassi, C. and G. Tate (2012). External networking and internal firm governance. *The Journal of Finance 67*(1), 153–194.

Friedkin, N. E. (1983). Horizons of observability and limits of informal control in organizations. *Social Forces 62*(1), 54–77.

Frydman, C. and D. Jenter (2010). Ceo compensation. *Annu. Rev. Financ. Econ. 2*(1), 75–102.

Gabaix, X. and A. Landier (2008). Why has ceo pay increased so much? *The Quarterly Journal of Economics 123*(1), 49–100.

Galenianos, M. (2014). Hiring through referrals. *Journal of Economic Theory 152*, 304–323.

Gayle, G.-L., L. Golan, and R. A. Miller (2015). Promotion, turnover, and compensation in the executive labor market. *Econometrica 83*(6), 2293–2369.

Gayle, G.-L., C. Li, and R. A. Miller (2016). Was sarbanes-oxley costly? evidence from optimal contracting on ceo compensation.

Granovetter, M. (1995). *Getting a job: A study of contacts and careers*. University of Chicago Press.

Granovetter, M. S. (1973). The strength of weak ties. *American journal of sociology 78*(6), 1360–1380.

Gregg, P. and J. Wadsworth (1996). How effective are state employment agencies? jobcentre use and job matching in britain. *Oxford Bulletin of Economics and Statistics 58*(3), 443–467.

Holzer, H. J. (1987). Hiring procedures in the firm: Their economic determinants and outcomes.

Holzer, H. J. (1988). Search method use by unemployed youth. *Journal of labor economics 6*(1), 1–20.

Huson, M. R., R. Parrino, and L. T. Starks (2001). Internal monitoring mechanisms and ceo turnover: A long-term perspective. *The Journal of Finance 56*(6), 2265–2297.

Hwang, B.-H. and S. Kim (2009). It pays to have friends. *Journal of financial economics 93*(1), 138–158.

Imbens, G. W. and T. Lancaster (1994). Combining micro and macro data in microeconometric models. *The Review of Economic Studies 61*(4), 655–680.

Ioannides, Y. M. and L. Datcher Loury (2004). Job information networks, neighborhood effects, and inequality. *Journal of economic literature 42*(4), 1056–1093.

Ioannides, Y. M. and A. R. Soetevent (2006). Wages and employment in a random social network with arbitrary degree distribution. *American Economic Review 96*(2), 270–274.

Jackson, M. O. (2010). An overview of social networks and economic applications. *The handbook of social economics 1*, 511–85.

Jackson, M. O., B. W. Rogers, and Y. Zenou (2017). The economic consequences of social-network structure. *Journal of Economic Literature 55*(1), 49–95.

Jog, V. and P.-L. Loh (2015). Information-theoretic bounds for exact recovery in weighted stochastic block models using the renyi divergence. *arXiv preprint arXiv:1509.06418*.

Kanade, V., E. Mossel, and T. Schramm (2016). Global and local information in clustering labeled block models. *IEEE Transactions on Information Theory 62*(10), 5906–5917.

Kaplan, S. N. and B. Minton (2006). How has ceo turnover changed? increasingly performance sensitive boards and increasingly uneasy ceos. Technical report, National Bureau of Economic Research.

Koller, D. and N. Friedman (2009). *Probabilistic Graphical Models: Principles And Techniques*. MIT press.

Krackhardt, D. (1987). Cognitive social structures. *Social networks 9*(2), 109–134.

Krackhardt, D. (2014). A preliminary look at accuracy in egonets. In *Contemporary Perspectives on Organizational Social Networks*, pp. 277–293. Emerald Group Publishing Limited.

Kuhnen, C. M. and J. Zwiebel (2008). Executive pay, hidden compensation and managerial entrenchment.

Larcker, D. F., E. C. So, and C. C. Wang (2013). Boardroom centrality and firm performance. *Journal of Accounting and Economics 55*(2-3), 225–250.

Leung, M. (2016). A weak law for moments of pairwise-stable networks. *Available at SSRN: https://ssrn.com/abstract=2663685*.

Maddala, G. S. (1971). The use of variance components models in pooling cross section and time series data. *Econometrica: Journal of the Econometric Society*, 341–358.

Marsden, P. V. and E. H. Gorman (2001). Interpersonal ties, social capital, and employer staffing practices. *Social capital: Theory and research*, 105–26.

McCarty, N., K. T. Poole, and H. Rosenthal (2006). *Polarized America: The Dance Of Ideology And Unequal Riches.* MIT Press.

McCarty, N. M. and K. T. Poole (1998). An empirical spatial model of congressional campaigns. *Political Analysis 7*, 1–30.

McKay, A. (2008). A simple way of estimating interest group ideology. *Public Choice 136*(1), 69–86.

McKay, A. (2010). The effects of interest groups' ideology on their pac and lobbying expenditures. *Business and Politics 12*(2), 1–21.

Meghir, C., R. Narita, and J.-M. Robin (2015). Wages and informality in developing countries. *American Economic Review 105*(4), 1509–46.

Miller, S. R. and J. E. Rosenbaum (1997). Hiring in a hobbesian world: Social infrastructure and employers' use of information. *Work and Occupations 24*(4), 498–523.

Montgomery, J. D. (1991). Social networks and labor-market outcomes: Toward an economic analysis. *The American economic review 81*(5), 1408–1418.

Montgomery, J. D. (1992). Job search and network composition: Implications of the strength-of-weak-ties hypothesis. *American Sociological Review*, 586–596.

Mortensen, D. T. and T. Vishwanath (1994). Personal contacts and earnings: It is who you know! *Labour economics 1*(2), 187–201.

Mossel, E., J. Neeman, and A. Sly (2014). Consistency thresholds for binary symmetric block models. *arXiv preprint arXiv:1407.1591*.

Neckerman, K. M. and J. Kirschenman (1991). Hiring strategies, racial bias, and inner-city workers. *Social Problems 38*(4), 433–447.

Newman, M. E. (2006). Modularity and community structure in networks. *Proceedings of the national academy of sciences 103*(23), 8577–8582.

Newman, M. E. (2013). Spectral methods for community detection and graph partitioning. *Physical Review E 88*(4), 042822.

Nowicki, K. and T. A. B. Snijders (2001). Estimation and prediction for stochastic blockstructures. *Journal of the American Statistical Association 96*(455), 1077–1087.

Pellizzari, M. (2010). Do friends and relatives really help in getting a good job? *ILR Review 63*(3), 494–510.

Peng, L. and L. Carvalho (2015a). Bayesian degree-corrected stochastic block models for community detection. *arXiv preprint arXiv:1309.4796*.

Peng, L. and L. E. Carvalho (2015b). Group-corrected stochastic blockmodels for community detection on large-scale networks.

Petrin, A. (2002). Quantifying the benefits of new products: The case of the minivan. *Journal of political Economy 110*(4), 705–729.

Poole, K. T. and H. Rosenthal (1985). A spatial model for legislative roll call analysis. *American Journal of Political Science*, 357–384.

Postel-Vinay, F. and J.-M. Robin (2002). Equilibrium wage dispersion with worker and employer heterogeneity. *Econometrica 70*(6), 2295–2350.

Rees, A. (1966). Information networks in labor markets. *The American Economic Review 56*(1/2), 559–566.

Rogerson, R., R. Shimer, and R. Wright (2005). Search-theoretic models of the labor market: A survey. *Journal of economic literature 43*(4), 959–988.

Ruan, H. (2017). Social capital and innovation: Evidence from connected holdings.

Rubinstein, A. (1982). Perfect equilibrium in a bargaining model. *Econometrica: Journal of the Econometric Society*, 97–109.

Schmutte, I. M. (2014). Job referral networks and the determination of earnings in local labor markets. *Journal of Labor Economics 33*(1), 1–32.

Shephard, A. (2017). Equilibrium search and tax credit reform. *International Economic Review 58*(4), 1047–1088.

Shue, K. (2013). Executive networks and firm policies: Evidence from the random assignment of mba peers. *The Review of Financial Studies 26*(6), 1401–1442.

Snijders, T. A. and K. Nowicki (1997). Estimation and prediction for stochastic blockmodels for graphs with latent block structure. *Journal of Classification 14*(1), 75–100.

Steger, A. and N. C. Wormald (1999). Generating random regular graphs quickly. *Combinatorics, Probability and Computing 8*(4), 377–396.

Stoer, M. and F. Wagner (1997). A simple min-cut algorithm. *Journal of the ACM (JACM) 44*(4), 585–591.

Taylor, L. A. (2010). Why are ceos rarely fired? evidence from structural estimation. *The Journal of Finance 65*(6), 2051–2087.

Taylor, L. A. (2013). Ceo wage dynamics: Estimates from a learning model. *Journal of Financial Economics 108*(1), 79–98.

Terviö, M. (2008). The difference that ceos make: An assignment model approach. *The American Economic Review*, 642–668.

Topa, G. (2001). Social interactions, local spillovers and unemployment. *The Review of Economic Studies 68*(2), 261–295.

Topa, G. (2011). Labor markets and referrals. In *Handbook of social economics*, Volume 1, pp. 1193–1221. Elsevier.

Trebbi, F. and E. Weese (2015). Insurgency and small wars: Estimation of unobserved coalition structures. *NBER Working Paper* (w21202).

Van den Berg, G. J. and G. Ridder (1998). An empirical equilibrium search model of the labor market. *Econometrica*, 1183–1221.

Yun, S.-Y. and A. Proutiere (2016). Optimal cluster recovery in the labeled stochastic block model. *in Advances in Neural Information Processing Systems 29*, 965–973.